DOCUMENT RESUME

ED 208 683 FL 012 601

AUTHOR	Yorio, Carlos A., Ed.; And Others
TITLE	On TESOL '79: The Learner in Focus.
INSTITUTION	Teachers of English to Speakers of Other Languages.
PUB DATE	79
NOTE	366p.: Selected papers from the Annual Convention of Teachers of English to Speakers of Other Languages (13th, Boston, MA, February 27-March 4, 1979).
AVAILABLE FROM	TESOL, 202 D.C. Transit Building, Georgetown University, Washington, DC 20057 ($13.00).
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Achievement Tests; Bilingual Education; Bilingual Students; Cloze Procedure; Curriculum Design; *English (Second Language); English for Special Purposes; Individualized Instruction; *Instructional Materials; Intelligence Tests; Language Proficiency; Language Usage; Material Development; Morphology (Languages); Reading Comprehension; Second Language Instruction; Speech Skills; Standardized Tests; Student Placement; Syntax; Teacher Education; *Teaching Methods; Technical Writing; Test Bias; Testing; Writing (Composition)
IDENTIFIERS	Community Language Learning; English for Science and Technology; Indirect Speech; Politeness; Silent Way (Gattegno); Speech Acts; Suggestopedia

ABSTRACT

Selections include Strevens' discussion of teaching for different circumstances, Morley's discussion of new developments in teaching materials, Swain's overview of research in bilingual education, Shuy's, Oller's, Bernal's, and Rosier's respective discussions of test bias, and Rivers', Hines', Scovel's, and Taylor's explorations of the Silent Way, Suggestopedia, and Community Language Learning. The section on curriculum design implementation includes reports on language teaching materials by Candlin and Breen and developing language use formats by Jones. Papers on testing include explorations of bias in language placement examinations by Farhady, bias in reading comprehension tests by Mohan, and alternatives to and scoring procedures for use on cloze tests by Mullen and Alderson, respectively. Focus on new approaches to training ESL nonspecialists, ESL teachers, and TESL students is provided by papers by Cott and Dubin, Newfield and Webb, and Clarke and Seward, respectively. Papers focusing on TESOL research include discussions of learner social and linguistic behavior by Cathcart, Strong, and Fillmore, politeness by Scarcella and by Walters, indirect speech acts by Carrell, questions by Crymes and Potter, oral and written syntactic relationships by Vann, composition skills by Arthur, and the relationship between syntactic skill and writing quality by Kameen. (JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208683

On TESOL '79
The Learner in Focus

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

James E. Alatis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Edited by

Carlos A. Yorio

Kyle Perkins

and

Jacquelyn Schachter

FL 012 601

Selected papers from the Thirteenth Annual Convention of
Teachers of English to Speakers of Other Languages
Boston, Massachusetts

February 27—March 4, 1979

Teachers of English to Speakers of Other Languages
Washington, D.C.

Copyright © 1979 by the
Teachers of English to Speakers of Other Languages
Washington, D.C., U.S.A.

All rights reserved. Printed in the United States of America

Library of Congress Catalog Card No. 80-50151

Additional copies available from:

TESOL

455 Nevils Building
Georgetown University
Washington, DC 20057

Preface

This is the last ON TESOL volume of the 70's. During this decade the TESOL profession has seen profound changes which have affected all facets of our work. These new trends have made us look at the student and the teacher in a different light, from a different perspective. As is almost always the case in education, these changes have been gradual, and often subtle. Those of us who have been involved in TESOL for a number of years, however, look back at the first classes we taught, the first materials we used, the conferences we attended, the journals we read, and find that the entire focus has changed.

The 70's constitute an era of many questions and few answers. It is the questions, however, that have made our task fascinating and challenging. These questions have a common core: how do people *learn* second languages? We now feel that *what* we teach and *how* we teach should revolve around that still unanswered basic question. Consequently, most of us try to be non-dogmatic about matters of method and technique because we feel that until the answer is found, flexibility and open-mindedness are the only sensible alternative.

But teach we must, and while we wait for language learning research to provide us with the final solution, we, in the classroom, are also focusing on the learners. We carefully consider their needs and goals, their likes and dislikes, their strengths and weaknesses. As a result, second language teaching today is much more learner-centered than it was ten years ago.

ON TESOL 79 attempts to capture this shift in focus by concentrating on the main issues of this past decade: second language acquisition research, new trends in teaching and curriculum design, current concerns in testing, and novel approaches to teacher training.

Each one of the plenaries in Part I, presents a perspective of one of the major areas of concern. Peter Strevens shows the enormous diversification which has resulted from a narrower focus on the learner: ESL, EFL, ESD, ESP, EAP, EST, etc. Joan Morley discusses the "new frontier" in the development of teaching materials. Merrill Swain gives us an overview of research in bilingual education. Darlene Larson's panel (Roger Shuy, John Oller, Ernest Bernal, and Paul Rosier) deals with the important issue of bias in testing. Finally, Wilga Rivers, Mary Hines, Tom Scovel, and Barry Taylor explore three teaching methods which have come to our attention in the last decade: The Silent Way, Suggestopoedia, and Counseling Learning/Community Language Learning.

Part II, Curriculum Design and Implementation, deals with issues that bear directly upon teaching. Part III covers most of the testing issues that have concerned us this past decade: bias, cloze procedures, and the design of new tests. Part IV, Teacher Preparation, focuses on the teacher as learner, and shows the need for new approaches to training. Finally, Part V, Research, contains papers that explore several areas of interest to teachers, curriculum designers, materials developers, and researchers: communicative competence, discourse analysis, and research in the area of writing skills.

Space limitations have made it impossible for us to include more of the many excellent papers presented at the Boston Conference. We trust that this selection, however, reflects the concerns of the profession at this time of transition and change, and we hope that this volume will contribute to developments in the 80's that will answer some of the questions raised.

—*C.A.Y.*
K.P.
J.S.

Table of Contents

On TESOL '79:
The Learner in Focus

PREFACE iii

I. PLENARY ADDRESSES

Differences in teaching for different circumstances or the teacher as chameleon .... *Peter Strevens* 2
Materials development: The new frontier, not by chance but by design .... *Joan Morley* 12
Bilingual education: Research and its implications .... *Merrill Swain* 23

Testing Panel

What makes test a four letter word?: Viewpoints on bias in standardized testing .... *Darlene Larson* 34
What makes language test an expletive: Tests, test selection, and test use .... *Ernest M. Bernal, Jr.* 35
Intelligence, achievement, competency, or language? .... *John W. Oller, Jr.* 40
The role of standardized tests at the school level ...... *Paul Rosier* 55
Three kinds of bias in standardized testing .... *Roger W. Shuy* 60

Methods Panel

An exploration of new trends in second language teaching: The silent way, suggestopaedia, counseling learning-community language learning .... *Wilga M. Rivers* 67
Explorations of new trends in language learning: The silent way .... *Mary Hines* 71
A review of suggestopedy .... *Thomas Scovel* 75
Exploring community language learning .... *Barry P. Taylor* 80

II. CURRICULUM DESIGN AND IMPLEMENTATION

Evaluating, adapting and innovating language teaching materials .... *Christopher N. Candlin and Michael P. Breen* 86
Frequency: A criterion for syllabus development .... *Jani Turano-Perkins* 109

Developing a format for working on language use with language learners .... *Marilyn Martin-Jones* 117
ESL speaking skills in the elementary bilingual classroom .... *Anna Uhl Chamot* 131
Teaching ESL technical writing with the personalized system of instruction .... *Gayle L. Nelson and John B. Keenan* 142
Rhetorical competence and EST discourse ..... *George R. Hepworth* 148

III. TESTING ISSUES

Test bias in language placement examinations ..... *Hossein Farhady* 162
Cultural bias in reading comprehension tests .... *Bernard Mohan* 171
Evaluating bilingual competence: An experimental innovative test .... *J. Donald Bowen and Sandra Plann* 178
An alternative to the cloze test .... *Karen A. Mullen* 187
Scoring procedures for use on cloze tests .... *J. Charles Alderson* 193
Testing in an EFL/ESP context .... *George E. Scholz and Celeste M. Scholz* 206

IV. TEACHER PREPARATION

Turning elementary teachers into ESL specialists: A cinderella story? .... *Nancy R. Cott and Fraida Dubin* 233
ESL teacher training: The need for a complete curriculum .... *Jacqueline K. Neufeld and Marion R. Webb* 239
A model for training the non-traditional TESL student .... *Mark A. Clarke and Bernard Seward* 250

V. RESEARCH

The social and linguistic behavior of good language learners .... *R. W. Cathcart, M. A. Strong and L. W. Fillmore* 267
On speaking politely in a second language .... *Robin Scarcella* 275
The perception of politeness in English and Spanish .... *Joel Walters* 288
Indirect speech acts in ESL: Indirect Answers .... *Patricia L. Carrell* 297
Questions in task-oriented discourse .... *Ruth Crymes and William Potter* 308
Oral and written syntactic relationships in second language learning .... *Roberta J. Vann* 322
Short-term changes in EFL composition skills ...... *Bradford Arthur* 330
Syntactic skill and ESL writing quality .... *Patrick T. Kameen* 343

REFERENCES 351

Part I

Plenary Addresses

Differences in Teaching For Different Circumstances Or the Teacher as Chameleon

Peter Strevens

The Bell Educational Trust

The purpose of this paper is to support a basic hypothesis about effective language learning and teaching. The hypothesis is that to be effective in promoting learning, teaching must take account of a large number of variables. The topic will be divided into three sections: first, the way in which our professional perceptions of the variables in language teaching have changed over the years; second, an analysis of the variables that seem to operate; and third, some consequences of these changing perceptions and these variables.

1. Our changing perceptions of the variables

Very few of us see more than a small selection of the main variables—just those variables, in fact, which operate with obvious effect in our own particular circumstances. What is more, even if all teachers agreed in their understanding of those variables they could observe and identify—which they don't—it is only by pooling and comparing their observations and their responses that we gradually establish a full perspective of the complex activity we are all engaged in.

In this regard we have been, in the past twenty years, somewhat in the position of those Blind Men of Indostan in the poem by John Godfrey Saxe (an American poet of the early 19th century), who encountered an elephant for the first time.[1]

---

[1] THE BLIND MEN AND THE ELEPHANT

A Hindoo Fable

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind.

The *First* approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant
Is very like a wall!"

The *Second*, feeling of the tusk,
Cried, "Ho! what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"

The *Third* approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The *Fourth* reached out an eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
"'Tis clear enough the Elephant
Is very like a tree!"

The *Fifth*, who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an Elephant
Is very like a fan!"

We perhaps are in a much happier position than Saxe's wise men because our 'elephant' has been reported, observed, photographed and subjected to minute analysis and discussion by annual Conventions for the past dozen years. As a result, we share a more comprehensive and accurate understanding of what an elephant is like than Saxe's blind Indostanis could muster—though at the same time we are increasingly aware that there is not just one elephant, but whole families and populations and even sub-species of elephants, and that therefore the scope of our study gets ever wider, even as our comprehension of it grows.

You will recall that one reason for the complementary analyses given by Saxe's blind men was that they each observed the elephant from a different standpoint. It is interesting to notice that a whole sequence of different viewpoints about the nature of our own profession, TESOL, is embodied in the sequence of acronyms or initials that have been employed on both sides of the Atlantic. The proto-TEFL term, as it were, was probably *ELT* (English language teaching), the original undifferentiated term used in Britain as early as 1945 and still employed today, for example in the title of the British *ELT Journal*, founded in 1948. Before long it became necessary, especially in British ELT, to distinguish between ELT in circumstances where the language had a special historical status in the community, for instance, English in Nigeria, or Hong Kong, or Fiji—(where it is referred to as a *second* language, hence *ESL*), and on the other hand ELT in circumstances where the language has no special standing and is not in widespread use—(where it is a *foreign* language), as in Japan, or Brazil: hence *EFL*. In British usage, when referring to the teaching of English to foreign students visiting Britain, the term *TEFL* became universal because such students almost always come from and return to 'foreign language' countries. To sum up the distinction, for us in Britain ESL indicates sociolinguistic conditions in a foreign country which compel significant changes in teaching English there.

In the United States, though, I sense a great deal of free variation, uncertainty, even change of meaning, in the use of the terms TEFL and TESL. For a given activity, sometimes the one is used and sometimes the other. There are perhaps two reasons for this: first, the American profession has been relatively little engaged in ESL in the British sense, i.e., in former British countries, so that this EFL/ESL distinction has had little relevance; second, the issue has been clouded by multiple meanings for the word 'second'. By talking of a

---

The *Sixth* no sooner had begun
  About the beast to grope,
Than, seizing on the swinging tail
  That fell within his scope,
"I see," quoth he, "the Elephant
  Is very like a rope!"

And so these men of Indostan
  Disputed loud and long,
Each in his own opinion
  Exceeding stiff and strong,
Though each was partly in the right,
  And all were in the wrong!

Moral

So oft in theologic wars,
  The disputants, I ween,
Rail on in utter ignorance
  Of what each other mean,
And prate about an Elephant
  Not one of them has seen!

John Godfrey Saxe, 1816-1887

'second language'—meaning the next one learned after the mother tongue—and by talking of a 'second foreign language' (or a third, or fourth)—for example, Spanish, French, or German in the school system—we have encouraged semantic ambiguity. By contrast, the name of our organisation, TESOL, is clear and unambiguous in the way it unites all branches of the profession.

What terms came next? There is some uncertainty about the chronology. My own understanding is that the next development occurred when the special and contentious problem arose of teaching English to children of the black community in the United States: was it to be regarded as 'mother-tongue' teaching? Should it be handled as a branch of TEFL? Black English became accepted as a dialect of English to be reckoned with in educational terms, and in this way *ESOD* or *TESOD* were born—and even *ESOLD* and *TESOLD*: teaching English to speakers of other languages and dialects.

Yet that is not by any means the end of these subdivisions within our own profession. The term *ESP* (English for specific purposes) has come into universal use to designate the teaching of English with particular restrictions on its aims, content or skill objectives. And within ESP, *EST* refers to English for science and technology, a particular sub-set of ESP which entails special learning features (and special teaching requirements, too) since the scientist or engineer has to learn 'the language of science'—whatever that is. The British Council favours a further internal distinction within ESP: *EAP* (English for *academic* purposes) and *EOP* (for *occupational* purposes).

There have been other TEFL-terms: *EIAL* (English as an international auxiliary language) had a brief existence in Hawaii before being dropped in favour of an INTER/INTRA distinction (English for *inter*national purposes, as in Japan, Brazil, etc. and English for *intra*national purposes, as for large populations of people in India, Malaysia and Singapore, and increasingly in other historically English-using countries).[2]

Two other terms now in constant use have not been turned into acronyms or initials. In the United States, 'Bilingual Education' is a specialised and established branch of English teaching, while in Britain we refer to 'Immigrant Teaching', which is that special branch of teaching English to the children of immigrants within the framework of the ordinary State school system—in fact, as a sort of TEFL-flavoured mother-tongue programe.

The lesson to be learned from this sequence of TEFL terms is that in the short space of twenty-five years our perceptions of some major variables within our own profession have repeatedly grown sharper and more delicately differentiated so that our existing ways of referring to what we do have become inadequate and have required the acceptance of a dozen or more specialised terms. And there will undoubtedly be more distinctions made in the future.

Yet the terms we have discussed above are only macro-variables. Beyond these macro-variables are a whole range of other choices, distinctions, alternatives, divergences, and these are variables whose operation affects our daily professional activity.

---

[2] See particularly Smith (1978) and Strevens (forthcoming).

2. An analysis of the variables[3]

When we attempt a reasonably comprehensive survey of the many variables which in total affect our profession, some teachers will already be aware of all the factors I shall mention, and more; to them, my apologies. Others, especially teachers who have always worked in one particular branch of TEFL, will be like Saxe's Blind Men of Indostan because they will have seen some of the variables in action while being unaware, perhaps, of others; teachers with different experience will recognise a different selection, depending on which part cf the TEFL elephant's anatomy they have been concerned with.

TEFL teachers who move from one kind of teaching to another,—from teaching 'general' TEFL to adults in the United States or Britain, to teaching children in Colombia, to teaching intensive ESP for medical staff in a country on the Arabian Gulf, to university TEFL in Scandinavia, to teacher training in Singapore—quickly discover that there are indeed a great many variables at work. Let us look at five types:

1. *Some of these variables appear as restrictions upon the teachers are actually permitted to do*: for example, in cultural matters. In some societies, for instance, the kind of boy-girl relationships commonly found in TEFL textbooks, illustrations of girls in short dresses, reference to alcoholic drinks, mention of dogs as domestic pets, even relaxed relations between teachers and students if they are of different sexes all may be culturally unacceptable.
2. *Other variables will appear as limits on what is physically and organisationally possible*: very large classes; classroom furniture bolted to the floor (which severely restricts communicative activity and work in groups); inadequate numbers of classes per week; a school year that is nominally of 30 teaching weeks but in practice may be of 20 weeks or less; lack of central authority or such draconian central authority that any deviation from normal practice, even creative experiment by intelligent teachers, becomes almost a criminal offence; absence of suitable textbooks, teaching aids or equipment, even of blackboards and exercise books.
3. Some variables will affect the teacher through *standards of teacher training* where the national average level of professional competence has pervasive consequences on what can be achieved in the classroom.
4. Yet other variables relate to *sociolinguistic attitudes and expectations*. If English is an unpopular language in a given place (e.g. in rural Quebec at present), teaching it may be ineffective for the learner and unpleasant for the teacher; if a community normally expects that its citizens will learn

---

[3] The identification of variables has a long history. The analysis in this paper is concerned more with emphasising the scope and diversity of types of variables rather than with comprehensivity in detail. It builds on the work of very many specialists. But one work above should be singled out for its subtlety and wide coverage of variables in the learning and teaching of languages, i.e. Mackey (1965).

English (e.g. in Holland, or Scandinavia) they generally do so, but conversely, low expectations (as of learning foreign languages by English school children) are fulfilled by generally poor standards of achievement.

5. *The educational framework* of TEFL, too, may vary. In some places the teaching of English begins at age 7 or 8 and continues as a part of normal, liberal arts education, for up to 10 or 11 years; elsewhere it starts only at age 14 and lasts for 4 years. By contrast, the framework for ESP is almost always that of adult education. It is often independent of the rigid administrative matrix of general education and is usually open for methodological originality—i.e. ESP is not tied to any particular method or materials. As a result the successes and failures in ESP are more directly caused by the teachers, as distinct from the system, than is the case in a general educational framework where the system often casts the variables into a mould which makes success difficult to achieve, save by exceptional teachers.

Five sets of variables have been mentioned so far: restrictions on teaching, physical and organisational constraints, standards of teacher training, sociolinguistic attitudes, the educational framework. They can all be roughly classified as derived from the community, that is to say, from the mixture of sociolinguistic conditions and the current realities of educational administration. Teachers in Britain or America or Australia are of course equally subject to constraints of this type, but in those countries teachers tend to be unaware of them and only realize their existence on first working abroad, or occasionally at home when some really major reform is being fought through our various pieces of legislative and administrative machinery.

From variables ascribable to the community and generally acting as constraints on the teacher, let us turn to other variables more directly under the control of teachers. It is necessary to do this in order to remind ourselves that although teachers work to a large extent within a matrix of conditions that are not of their choosing and are only minimally under their control, they do have the responsibility for selecting among a wide range of further variables.

The most obvious sets of teacher-controlled variables fall under the twin headings of *syllabus* and *methodology*, i.e. the large and growing range of instructional techniques available to the teacher (methodology) and the principles for selecting and organising the content to be taught (syllabus design).

The two are inter-linked. Modern views about the importance of the learner lead us to analyse the learner's needs in ever-greater detail, to establish with maximum certainty his precise aims and objectives; to specify the language and other content he will require; to organise the sequence in which it is most appropriate to teach this content to this learner; to consider what teaching techniques can most effectively be used by the teachers available; and to teach this content in this sequence to these learners with these needs.

The first two parts of that process, analysing the learner's needs and objectives and determining appropriate content and sequence, form part of *syllabus design*. This is a process where members of the teaching profession consider several dimensions in order to establish the best possible 'fit' or matching. What are these dimensions? Here are some of them: the identity of the learners and their principal characteristics (age, educational level, stage of proficiency already reached, etc.); their aims and objectives (to pass a particular examination, to achieve practical oral communication with certain types of people, to read with understanding certain kinds of written text, etc.); the language appropriate to those objectives (which may either be 'general English' or may be some defined sub-set of English, like 'English for Indian village tropical health projects', or 'English for air traffic control', etc.) any special need for some items to be ordered before other items (e.g. in an ESP course for ships' officers, to teach compass directions and bearings early on, where in contrast they might be taught late or never in a course for specialists in tropical medicine).

The syllabus designer will bear in mind several different aspects of the content: linguistic, situational, notional, functional and communicative. *Linguistic* aspects: the language content to be learned in terms of grammar, vocabulary, pronunciation, formulae and fixed expressions, idioms, and so forth; *situational* aspects: how to make the language content more interesting, more easily learned and remembered, and more like real life by teaching it in relation to familiar or imaginative situations, like at the Post Office, telling the time, going to school, and so forth; *notional* aspects: ways in which various fundamental notions about the universe are expressed in English notions of time, of place, even of case, (i.e. who does what, who to, and with what result); *functional* aspects: how English expresses certain functions of language, such as negation, possibility and impossibility, description, questioning, judgements, and many others; and *communicative* aspects: how to use English for meaningful, deliberate, effective communication between human begins, including discoursal and rhetorical rules.

Methodology, too, is a variable whose manipulation is normally open to the teacher, not imposed upon him or her by outside authority. The term includes both strategic and tactical decisions. Strategic decisions include such questions as: the balance of importance and time to be given to the various skills of speaking, reading etc.; the extent to which overt use is made of grammar, and what kind of grammar it will be; the extent of individualised procedures, if any, the relation between the teaching syllabus and any terminal assessment or examination; the degree of reliance upon class work and private study; the use of language labs and other aids; and many others. Tactical decisions include: the selection of particular teaching techniques at particular points in each course and each lesson; the flexibility of teachers to change their techniques from moment to moment according to the learning paths of the students; the encouragement of methodological originality and creativity on the part of the teacher (and even the learner, on occasion); and many more choices of this limited kind.

Compared with the central importance of syllabus design and methodology, the variables of *materials production* and *evaluation* are perhaps secondary, even though their impact on the daily classroom life of the teacher and the learner is often considerable. There is, however, one further variable on the teaching side (as distinct from the community side) which must not be overlooked: *teacher training*. To the individual teacher, teacher training is usually looked at in terms of preparation for his or her personal career. The individual learner rarely even considers the question, except perhaps to reassure himself that he is being helped to learn by someone who is at last not an amateur—much as the patient awaiting surgical treatment likes to be certain that he is not in the hands of a do-it-yourself hobbyist. Yet on a larger scale the extent, nature and quality of the teacher training available in a given country will crucially affect the quality of teaching that is normally given there; and it will also largely determine which choices are made under the variables of syllabus design and methodology. So when we observe different kinds of teaching and different standards of teaching taking place under different conditions we can be sure that the kind of teacher training undergone by the teacher will be a major determiner thereof; and we can also be certain that if we are seeking ways of modifying the teaching and learning in a given place, we shall be bound to include the teacher training system in our calculations.

The variable of teacher training embraces many elements. If one accepts that teachers are by definition members of the educated sector of the community (and that the notion of an uneducated teacher is, or ought to be, a logical contradiction) then a prime element in teacher training concerns the level of personal education possessed by the teacher trainee. A case can be made, for the view that teaching, being a matter of close and continuous interaction between individual human personalities, also requires a high level of emotional maturity and stability. Then one must consider the balance between the three main components of a teacher training course: a *skills* component which develops practical, instructional techniques, both those common to all branches of teaching and those that are special to TEFL, including an adequate command of the language he or she is teaching; an *information* component, in which the teacher takes in the very considerable body of knowledge about education, teaching, language, English today, sociology, psychology, the organisational framework s/he is working in, and much more; and in addition a *theory* component, which provides him/her with an intellectual basis for knowing not just *what* to teach and *how* to teach it, but also *why* to teach that rather than something else.

Let me offer just one example of the kind of information that ought to be included in TEFL training courses. It concerns the astonishing changes that are taking place in the spread and functions of English on a global scale. Young teachers need to be informed about the way that English is now widely regarded, abroad, as an international possession; it is no longer the cultural property of the British, the Americans, the Australians and New Zealanders. Teachers need to realize that in a growing number of countries e.g. India, Singapore, parts of Africa, English is used by vast and growing populations of people (28 millions in India alone) who never meet us native speakers and who have no desire or need to model their English on ours. We have to get rid of our monolingual ethnocentricity and accept the existence of a great many localised forms of English, one characteristic of which is precisely that they are different from British and Australian and American English. There are many other kinds of information to be included: this may serve simply as an example. And yet, as we shall see later, even an ideal teacher training programme supplied to every TEFL teacher at the outset of his or her career fails to meet all the needs of the profession.

We must now look briefly at one other set of variables that affect our teaching: variables relating to the *learner*. To list them all would be a major task. We now recognise that TEFL learners are human beings (as distinct from laboratory animals) and that while they share certain universal human characteristics they also each carry a wide range of strictly individual features. How can we distinguish those that are significant from among the mass of learner characteristics?

There are three learner variables above all, it seems to me, which lie at the heart of the interaction between learner and teacher and determine how effectively the learner learns. These are: the learner's *reasons for learning*, his or her *attitudes towards learning*, and his or her *expectations of learning*.[4]

Learners' *reasons* for TEFL learning vary very greatly. For young children, it is normal and enjoyable to do what teacher says; for adolescents reasons for TEFL learning are often founded on hazy notions of 'relevance' and on emotions of love or hate towards the teacher, and they are frequently negative reasons, so that learning does not occur; for adults, the reasons are generally quite definite and instrumental.

Learners' *attitudes* towards TEFL learning are a compound of their attitudes towards learning in general, towards English in particular, towards the teacher and the textbook, and even towards themselves as learners.

Learners' *expectations* of TEFL learning are based partly on their past history and experience of language learning, partly on opinions fed to them by their family, their friends and their teacher, and partly on the current folk-myth about whether his/her community normally expects people to learn English.

As far as teachers are concerned, they quickly discover two things: first, the success of their own efforts in the management of learning and of their learners' efforts in achieving their optimum learning rate (Strevens 1977b), depends heavily on these reasons, attitudes and expectations. Second, much of their time and thought is devoted to unobtrusive efforts to make these factors as positive as can be achieved.

---

[4] I am deliberately avoiding the term *motication*, partly because it has different meanings for different people, but chiefly because motivation embraces a large number of factors. The 3-term analysis into reasons, attitudes and expectations is made from the standpoint of the teacher rather than on a basis of psychological or sociological analyses.

To sum up this penultimate section, then, within the totality of TEFL there exists a large number of variables, identifiable broadly as relating to the community, the teaching profession itself, and the learner. In addition, the nature of the profession's perceptions about its own task changes and develops gradually with the passing of time.

In short, the teacher is like a chameleon. To be efficient, i.e. to survive, the chameleon must continually observe his surroundings and adapt to them. And adapting means action. It means selecting from among a range of possible colourings and patterns precisely those which are appropriate to the moment. The teacher, to be efficient, perhaps even to survive, needs to adapt in a similar way: to select from among a wide range of possible techniques and courses of action precisely those which are appropriate to the circumstances of the learning teaching situation. In the third and final section of my paper I shall comment on the responsibilities of the teacher and of the profession.

Teaching is a hard and emotionally bruising occupation, and language teaching more so than most other specialisms, especially if the language one is teaching is not one's mother tongue. Teachers essentially need two distinct kinds of follow-up to be available to them after and beyond their initial training. One kind of follow-up must be the career-long provision of morale-boosters —of events where teachers can meet together, share ideas, renew acquaintance with success, seek professional commiseration with our little deprivations, and generally break free of the frustrations and annoyances of the daily round (the TESOL Convention is the outstanding example of this kind of event); also teachers' associations, workshops, conferences, magazines and journals, etc., and, for teachers for whom English is not the mother tongue, opportunities to restore and improve their command of the language.

The second kind of follow-up is of a different kind and affects only a proportion of teachers, though again it is essential to the onward development of the profession as a whole. This is the availability of courses of further training, with the purpose of developing an adequate higher echelon of senior professionals, teacher trainers, advisers, and consultants.

3. Consequences of this analysis

The argument so far presented is basically a simple one. It is that the profession we belong to is now a great and global one, of vast scope and influence and containing enormous complexities and subtleties. No single one of us, as a teacher, encounters all the possibilities of choice. Indeed, most of us live our working lives within a single, fairly simple segment of the whole; we are usually aware of only the big toe of the TEFL elephant, or the tip of its tusk, or its tail, or its trunk. But just as the elephant exists as a complete and complex animal even if we can see only a part of it, so also TEFL exists, and is subject to all the many variables we have been discussing.

But we are teachers. We are members of the total profession of human education. And that places before us the responsibility for knowing what we are doing, for being aware not just of our own little segment of the profession, but (at least in broad outline) of the whole field, of the variables that now exist and of their consequences. What are these consequences?

The first and most obvious consequence is that any lingering nostalgia for a single 'best' method is completely mis-placed. No single method can conceivably be equally suitable for all values of all the variables, for all learners, of any age, regardless of aims, of attitudes, of level of proficiency, and so forth.[5]

A second consequence is that the individual teacher needs to acquire the widest and deepest understanding of all the variables he or she is likely to encounter in the language learning/teaching situation where s/he will be working. The teacher needs to select, devise and operate, for any given situation, that methodology which has the best fit, the closest match, with all the variables.

Behind the teacher is a third and final consequence of the variables facing us. This is a consequence for the TEFL profession. Now that TEFL is so vast, so complex, yet so important, it has become essential for the profession to have corporate ways of observing, monitoring, analysing and influencing its own development. Hence the vital importance not only of academic centres of excellence such as university departments and teacher training colleges, not just of resource centres like the Centre for Applied Linguistics, the Regional English Language Centre in Singapore, the British Council's English Teaching Information Centre in London, and similar centres elsewhere, but also of professional teachers' organisations—of TESOL, IATEFL, AILA and the rest—and of their programmes of activities.

This, then, is the link between the relatively drab practical daily activity of our working lives, and the brightly-coloured programme of a TESOL Convention. Here is where we learn from each other about variables other than the ones we know of and meet in our ordinary life; here is where we exchange experiences, and where we discover that the wall, the spear, the snake, the tree, the fan and the rope, are in fact part of an elephant. Our profession becomes daily even more complex, subtle, sophisticated. We as teachers must learn to adapt, chameleon-like, to an ever-greater array of variables, so that we can offer to our students not a single technique which may or may not be effective, but the best possible choice of teaching for the particular variables that operate in our own students' individual circumstances.

---

[5] If I have any reservation of substance about the claims made for the Silent Way, Community Language Learning and Suggestopedia, it is that they seem to claim to be suitable in all circumstances. I find that to be counter-intuitive, and against experience.

Materials Development: The New Frontier, Not By Chance But By Design

Joan Morley

The University of Michigan

1. Introduction

Materials development in ESL has many dimensions.[1] Reflect for a moment on the following concerns expressed by respected colleagues. Each addresses a different aspect of the art/science of developing instructional materials for second language learners. In later sections of the paper special reference will be made to each of these comments.

> The higher the general standard of the teachers, the less important are the course-books and other teaching materials. But when the standard of teachers is low, as it inevitably is for a time in conditions of rapid expansion of education, then the standard of teaching materials assumes great importance. This is what is meant by 'the teachers/materials equation.' Insufficiently-trained teachers working with poor materials face a considerable handicap. When this occurs it is a defect of the system, for whose remedy a three-way collaboration is necessary, between publishers, administrators, and teachers. (Strevens, 1977a)
>
> If authors of textbooks of the future are to respond to all of the significant changes taking place in language pedagogy, they cannot overlook the social climate in which learning takes place. In humanistic classrooms—and there are more and more of them—people take responsibility for their own learning. (Dubin, 1978)
>
> Efficient language teaching must work with, rather than against, natural processes, facilitate and expedite rather than impede learning. Teachers and teaching materials must adapt to the learner rather than vice versa. (Corder, 1976)
>
> Every teacher is an adapter of the text used. The more experienced and sensitive the teacher, normally the greater the adaptation. We're adapting even when we refer to an exercise covered earlier or when we complement the lesson with realia, music, or unprescribed student activities. (Madsen and Bowen, 1978)
>
> . . . adaptation is invitable; it ought therefore to receive more attention and more prestige than it usually does. . . . language study is inevitably a total human experience; writers and teachers ought therefore to act as though it is. (Stevick, 1971)

2. Significant developments affecting ESL

Twelve years ago TESOL was established as an independent organization dedicated to the professional study of teaching English to speakers of other languages and dialects. Since that time many significant developments have taken place in the field of second language learning and teaching in related disciplines in world affairs.

---

[1] The phrase "materials development" is to be interpreted with the broadest of definitions throughout the paper unless otherwise specified.

In the following section three specific developments will be discussed. Each has had a particularly strong impact on the field of ESL. Their direct influence on materials development in ESL will be examined.

1. *The ESL "population explosion"*. Within the current world population of around four and a quarter billion people it is reported that around 374 million speak English as their primary language. Added to this, and of particular concern to TESOL, an estimated 305 million persons use English as their secondary language. This number continues to grow yearly, as does the growth-rate figure. The result, world-wide, is a burgeoning "population explosion" of ESL students—students who need effective instructional programs carried out by well-trained ESL teachers.

Decker (1978) has compiled the following information on English-language use: (1) English is the most commonly taught foreign language in the schools of countries where it is not the mother tongue of the inhabitants, including the U.S.S.R. and the People's Republic of China, (2) in the most significant technical and critical fields—science, education, commerce and trade, technology, medicine, engineering, international politics, and journalism—there are more publications in English than in any other language, (3) more books, newspapers and magazines, radio and television programs, records and tapes, and films are produced in English—and distributed more widely—than in any other single language, (4) English is the most commonly used language at international conferences, (5) English is the official language employed in international air traffic and in international shipping, (6) three-quarters of the world's mail is written in English.

While estimates of the numbers of persons who use English as their secondary language are well over 300 million, often documented information on actual numbers of students who make up individual ESL populations is difficult to obtain. In the United States, in fact, the National Center for Education Survey (NCES) only recently completed studies which provided data on the numbers of learners who make up various segments of the ESL population, on the numbers of teachers involved, and on the nature of teacher qualifications.

The NCES survey supplied significant and sobering information in three areas which are of particular concern to ESL professionals: (1) a totaling and subtotaling of the enormous numbers of present/potential ESL learners, (2) an accounting of the alarming deficit in the number of minimally trained (let alone well-trained) ESL teachers available, and (3) a specification of gross inadequacies in the teacher-preparation of teachers who are pressed into service as ESL instructors. Dorothy Waggoner, (1978) Education Program Specialist for NCES, reported the following survey information:

> An estimated 28 million persons in the United States (one in eight) have non-English language backgrounds, and contrary to general belief, most are native-born, not foreign-born.

> Over 5 million, of the 28 million who have non-English language backgrounds, are of school age, between the ages of six and nineteen—over ten percent of the persons in this age group.
>
> An estimated 2.4 million persons, aged four and older, *do not speak English at all.*

Waggoner also reported the following mind-boggling gap between teacher need and teacher preparedness:

> In the public schools in the United States during the school year 1976-1977, out of an estimated one hundred and two thousand teachers (102,000) actively engaged in teaching English as a second language, only THREE out of TEN had taken even one course—one course—in teaching English as a second language.

Peter Strevens' (1977a) observations are directly and disturbingly applicable here: "... when the standard of teachers is low, as it inevitably is for a time in conditions of rapid expansion of education, then the standards of the teaching materials assumes great importance. Insufficiently-trained teachers working with poor materials face a considerable handicap."

The serious state of affairs revealed by the NCES survey is a clear instance, indeed, over 70 thousand instances where the quality of the learning materials assumes critical importance. Finally, as the number of ESL learners increases and the need for more and better prepared ESL teachers intensifies, there is an additional feature of ESL population growth which must be taken into serious account—the special purpose English-language needs of students in an increasing variety of particular circumstances. These needs must be analyzed and appropriate learning objectives formulated in order that the most useful instructional programs and learning materials can be designed. Clearly, the challenge to the ESL profession is a demanding one which continues to expand into new dimensions. Our responsibility is, as G. Richard Tucker's closing plenary statement charges (1977), "... to provide effective second language training to an increasingly large and heterogeneous group of students".

In relation to the topic of this plenary session then, the need to provide not just more, but high quality and increasingly specialized learning materials is a central responsibility of our profession and must become a priority concern.

In many ESL learning/teaching settings around the world the written texts and recorded tapes exert a powerful and unquestioned force. This is especially true whenever and wherever teachers of English as a second language have had neither broad study of language and language-learning, nor thorough training in ESL. Indeed, in the United States alone, if we consider the serious impact of the one Waggoner figure (i.e. of 102 thousand teachers assigned to teach ESL, seven out of ten are totally untrained in ESL) then in the classrooms of over 70 thousand teachers who have never taken even one course in ESL, the textbook may well be responsible for teaching the teacher as well as the student.

2. *The "knowledge explosion"*. During the past two decades there has been an extensive and wide-ranging increase in professional study relating to theory and research in the social and behavioral sciences. Some have called it a "knowledge explosion".

Developments in psychology and linguistics have been of major importance to the field of second language learning and teaching, especially those in the emerging interdisciplinary fields of sociolinguistics, psycholinguistics, and neurolinguistics and particularly, studies in first language acquisition.

Beginning in the late 1960's and continuing to the present time, there has been a smaller explosion of information, but a significant and respected one, related directly to theory and research in second language acquisition. Increasing numbers of conferences papers, journal articles and new books appearing on publishers' inventories are a measure of the emerging interest in this field of specialized study.

A survey of TESOL convention programs reveals the growing prominence of second language acquisition study within our own organization. Compare: Ten years ago, in 1969, at the third convention of TESOL, (Chicago), out of nine pre-convention study groups, and 83 convention presentations, one paper dealt with a single facet of second language acquisition theory and research. Four years ago, in 1975, at the ninth annual convention, (Los Angeles), one out of 36 pre-convention sessions and 22 out of 161 convention papers dealt with a variety of facets of second language acquisition. And in 1979, at the 13th convention, (Boston), out of a total of approximately 300 pre-convention and convention presentations, over 50 address diverse topics in an ever-widening range of studies in second language research.

As theory and research in second language learning and teaching have opened up new vistas it has been tempting to look for immediate applications in the classroom. Indeed there is a potential two-way street of interchange between theory and practice which will benefit both but observe that both directions are well-marked with signs for caution. One direction runs from developments in theory and research to changes in classroom instruction; the other, reciprocally, from changes in classroom instruction to further developments in theory and research. But let the traveler beware! Research scholars have been the first to stress the need for caution, lest the traversing of this two-way street be premature.

Elaine Tarone, Merrill Swain, and Ann Fathman, (1976), discussed *Some Limitations to the Classroom Applications of Current Second Language Acquisition Research*. They began with this assessment:

> Research in second language acquisition has developed rapidly in recent years. With this rapid development has come an increasing desire on the part of some teachers and researchers to apply the results of research to the language classroom. However, although progress has been made, the reality of the situation is that second language acquisition research is still in an infancy stage, and hence cannot yet provide the classroom teacher with the kind of reliable guidelines needed to effect curriculum change.

G. Richard Tucker (1977) reinforced their call for caution with these words:

> I would argue strongly (as do Tarone, Swain, and Fathman) that it is premature to consider revising syllabi or instructional approaches to reflect the results of these studies; although . . . there are certain broad areas of application for the results of second language research.

Finally J. P. B. Allen, (1976), has made the following observation:

> . . . there is perhaps, something wrong with the idea that the only way to 'apply' results of research is to write a whole new textbook or a brand new curriculum sequence. Perhaps it is better to see the current applications of research as comprising an influence which indirectly and subtly changes the teacher's attitude towards what she or he is trying to do in the classroom—changing the teacher's attitude towards errors, for example, or leading the teacher to pay more attention to forms the students are producing. Such a change in attitude may be the most important application of current research which can be made to the field of language teaching.

In relation to materials development then, it is essential for all ESL professionals to keep abreast of continuing developments in theory and research in language, language learning, and language behavior. From an informed position knowledgeable judgments relating theory and practice can be made and potential implications and applications, direct or indirect, explored or exploited.

3. *The "instructional revolution" in ESL.* A third development began in the late 1960's and quickly gained both momentum and magnitude during the 1970's. Called an ESL "revolution" by some it has brought about significant changes in the form and in the substance of ESL classroom practices. The emergence and rapid growth of this movement related primarily to two factors: (1) a growing disenchantment with traditional ESL methodologies and materials, and with student achievement, and (2) an increasing demonstration of the failure of structural linguistics and behavioral psychology to account for crucial properties of language, language learning, and language behavior.

A review of TESOL publications over the past twelve years clearly documents the state and the stages of change. Forty-eight *TESOL Quarterlies*, 62 *TESOL Newsletters*, and five *ON TESOL's* chronicle problems and solutions, conflicts and resolutions, accusations and refutations, attacks and defenses, paradoxes and panaceas, dysarthrias and dyspepsias.

As a profession, we have been engaged in re-evaluating and restructuring curricula, methods, materials. These endeavors have affected every single domain of TESOL—English as a second dialect, EFL for foreign students in English-speaking countries, teaching English abroad, and ESL in elementary education, ESL in secondary education, ESL in higher education, ESL in adult education, and ESL in bilingual education.

Significant features of the "instructional revolution" in ESL include the following: (1) a focus on the individuality of the learner, (2) a focus on the learner's language, (3) a focus on the special purpose needs of particular students, (4) a focus on genuine communication and (5) a focus on a humanistic classroom. Indeed, as Dubin (1978) noted in addressing the fifth of these features,

> If authors of textbooks of the future are to respond to all of the significant changes taking place in language pedagogy, they cannot overlook the social climate in which learning takes place. In humanistic classrooms—and there are more and more of them—people take responsibility for their own learning.

Perhaps the extent of changes in ESL classroom practice can best be captured in noting a recharacterization of the roles of teacher, learner, and learning materials—new sets of expectancies and new responsibilities.

3.1 **Teachers:** New insights have moved us to reject as untenable a concept of the teacher role as primarily one of model and drill-leader. The teacher role today emphasizes the importance of the teacher as a manager of learning experiences and a facilitator of learning. Teacher-training programs today and in-service sessions on teaching resemble very little those of ten or even five years ago.

3.2 **Learners:** Similarly, we no longer find tenable a concept of the student role as primarily that of repeater of forms and patterns. Today students are recognized as the active creative prime-movers in their own learning process. In Corder's (1978) words: "Efficient language teaching must work with, rather than against, natural processes, facilitate and expedite rather than impede learning. Teachers and teaching materials must adapt to the learner rather than vice versa."

3.3 **Learning Materials:** In ESL classrooms the written textbook has always been a central force. It dictates the specific written sample of English to which a group of students is exposed, and it controls the specific spoken subset of English which the teacher and the students use as they the textbook and engage in activities related to it. Said another way, the written textbook determines, in large part, the nature of the language input which is made available for students. It provides the language data from which students form and shape a view of target English.

By-and-large the days have passed in which dependence upon one book (or set of books) based on linguistic forms alone can be considered the most fruitful base for an efficient and effective job of facilitating second language learning. Today teachers' concerns about learning materials are influenced by changing assumptions about language, language learning, and learner processes. In particular, our view of the role of learning materials has expanded to include a central emphasis on communication and language function, as well as on language form.

In relation to materials development then, note the impact of these important changes in ESL concepts. *First* with classroom attention moving to a special emphasis on language function as well as on language form, finding an appropriate selection of texts becomes more difficult. *Second*, once selected, prepared textbooks almost always need to be adapted in order to provide for specific student needs. As Madsen and Bowen (1978) observed: "Every teacher is an adapter of the text used. The more experienced and sensitive the teacher, normally the greater the adaptation." *Third*, in addition, texts often need to be enriched with supplementary materials. Increasingly, teachers are expected to become their own part-time materials writers. This, in turn, places more pressure on college and university teacher-training programs to expand their course work in textbook selection and materials adaptation and to develop courses in classroom research and materials development. *Fourth*, with the demand for new texts on the increase, there is a more and more urgent need for creative, knowledgeable, experienced teachers to become full-time or at least part-time professional textbook writers.

In the larger picture, if the research and experimental work of materials development is to yield the highest quality product, more and more college and university departments and school systems, as well as government agencies, will need to do what a number already have done—support their commitment to advancing the efficiency and effectiveness of second language learning and teaching with reasonable financial subsidy and related teacher time.

3. Expansions in ESL

Three significant developments in ESL have been discussed, the ESL population explosion, the knowledge explosion, and the instructional revolution in ESL. Working together, these three forces have contributed to a period of enormous expansion in the teaching of English as a second language, an expansion which has been felt in every aspect of professional activity.

We need look no further than the growth of our own organization to see the impact of expansion. As Executive Secretary James Alatis (1976a) has summarized, ". . . once an organization goes beyond 5000 members, it is no longer a small organization, or even a medium-sized one. It is a large organization, and a powerful one." Today TESOL membership is over 8000; when TESOL began in 1967, the membership was 357.

In the following section three important expansions of the last twelve years will be cited.

1. *Professional education.* One expansion is the ways and means by which we deal with professional education. In teacher preparation, in order to keep pace with mounting demands, college and university training programs have increased in number and have expanded the scope of course offerings and practice teaching experiences. As teachers we feel an urgent need to keep up with the knowledge explosion and to monitor developments in the instructional revolution. To cope with this need, organizational forums have increased in number, in frequency, and in scope. We attend more and more meetings—large national and international conferences—and smaller ones under the auspices of national, regional, state, provincial, and city jurisdictions. Some of our forums are more specialized, and some are more comprehensive than ever before. (Compare, for example, the TESOL convention programs of 1969 and 1979. In 1969 the convention index included the names of 142 participants; in 1979 the convention index listed 877 names.) Similarly, in the area of in-service train-

ing, there is continual increase and diversification. In short, we get together more often, in more places, and have more to share with one another.

2. *Textbooks available*. A second expansion is apparent in the textbook marketplace. In the past twelve years, publishers' inventories have doubled, and in some cases tripled. More textbooks are available today including a wider variety of specialized types of learning materials which did not even exist a few years ago. Yet there is a constant pressure for new texts and materials, and hence a need for more teachers to make a part of their career option one of professional materials development.

3. *Teacher responsibilities*. A third expansion involves the increased responsibilities of ESL teachers. Despite our preoccupation with reassessments of classroom practices and despite emerging redefinitions of what language learning may actually entail, unfortunately, unlike industry, we cannot stop our work while we retool for instructional change. It is the ESL teacher in the classroom who must institute gradual but continuous change, on the one hand, while carrying on the details of daily operation, on the other.

As noted previously, even textbook selection is no longer necessarily an easy task. Further, once chosen, the prepared text usually must be adapted and often supplemented. As Stevick (1971) has noted in the introduction to *Adapting and Writing Language Lessons*, ". . . adaptation is inevitable; it ought therefore to receive more attention and more prestige than it usually does."

In short, a large part of the responsibility for instructional change is borne by the ESL teacher who must become teacher/curriculum writer and teacher/materials developer. For many teachers, however, these are jobs which they do not feel altogether well-equipped to do without further course work and/or special in-service training programs. Without the assistance of specialized training, many teachers who must assume the responsibilities of designing curriculum and developing materials, will continue to be frustrated by the disparity between the training available and the requirements of their jobs.

4. Specialized training in ESL materials development: a professional responsibility

Significant developments and expansions in ESL have been reviewed and related to materials development. In this section a call will be made for expanding professional training in aspects of evaluation/adaptation and research/development of ESL instructional materials—an increasingly critical area of ESL responsibility and one which must become a priority concern. Specific suggestions for expanding this component of ESL professional education will be outlined.

1. *Professional planning for materials development*. As a profession, what can we do that we are not now doing to best prepare to meet the following needs: (1) a growing demand upon ESL classroom teachers for more and more materials adaptation and supplementation, (2) an increasing need for more,

highest quality, both general and increasingly specialized textbooks and other learning materials.

Clearly, we are moving rapidly into new frontiers in materials development. The new frontiers are here and now, and they need the most thoughtful of explorations. There is a gap between theory and practice. But the gap will not be narrowed by fads, nor by putting new names to old theoretical and pedagogical friends, nor by jumping on bandwagons of premature attempts to translate theory into practice. Less than well-informed and well-formulated attempts to do so can only contribute to disappointments, to instability, and to an illusion of bridging the gap . . . which under close scrutiny is more apparent then real.

Materials development is a creative process, one which is exciting, challenging, time-consuming, exacting, and often frustrating. There is a built-in need for experimental trial-and-error, and no matter how careful the planning, some developments do occur by chance. But the highest quality, most efficient, and most effective learning materials are seldom the product of chance. They are the product of design. And quality design is the product of three fundamentals: (1) sound knowledge of theoretical constructs relative to the project, (2) sound knowledge of pedagogical constructs relative to the project, (3) experience with the project in ESL classrooms (continuing throughout every interim phase of development).

Further, as design is translated into substance, it is the responsibility of the materials developer to maintain a continuous monitoring of issues relating to language and language learning processes, to the language learner, and to the language teacher.

Again then, as a profession, . . . what can we do that we are not now doing to best prepare to meet the present and future needs in materials development in ESL? The answer necessarily lies in yet another expansion in professional education, in career preparation for future ESL teachers, in continuing education, in education for change, and in ourselves. In the three subsequent sections suggestions will be outlined for expanding specialized training in evaluating and adapting instructional materials and in researching and developing materials: in teacher training programs, in in-service training programs, and in convention programming.

2. *Teacher training programs. Suggested Curriculum Expansion:* An increase in specialized course offerings in ESL materials development with a two (or three) course sequence which provides for a balance of: (1) theoretical foundations and the form and function of learning materials; (2) practical experience in evaluating, adapting, researching, and developing materials. Course content possibilities include the following:

(1) a chronological survey of theory in second language learning and teaching; special attention to influences on curriculum design and the form and function of instructional materials.

(2) a review of the history and development of ESL learning/teaching materials; special attention to the needs of intended student audiences and to a study of geographical use patterns.

(3) individual and/or group projects in criterion-referenced evaluations of selected ESL textbooks (and other learning materials); special attention to goals, intended student audience, student involvement, content, sequencing, presentation format(s), measures for assessing learning, and teacher involvement, as they reflect theoretical/pedagogical considerations.

(4) individual and/or group projects in assessing the needs of a particular group of students, selecting a set of appropriate learning materials, outlining intended use schedule, and projecting possible adaptations and supplementations.

(5) specific individual and/or group projects in researching, designing, and developing a (limited) set of instructional materials.

3. *In-service training programs. Suggested Program Expansion:* An increase in specialized in-service training programs devoted to ESL materials evaluation/adaptation and research/development:

(1) under the individual or joint sponsorship of public or private school systems, community colleges, adult education programs, public or private institute programs and/or government programs, and

(2) focused on the relevant materials development needs of one or more (closely related) areas of TESOL: English as a second dialect, EFL for foreign students in English-speaking countries, teaching English abroad, ESL in elementary education, ESL in secondary education, ESL in higher education, ESL in adult education, and ESL in bilingual education.

Format and content possibilities (with attention to both theory and practical experience) include the following:

(1) mini-courses in reviewing recent developments in theory in ESL learning and teaching; special attention to implications for curriculum design and materials development for ESL students in specific educational settings.

(2) study groups in evaluating ESL materials; special attention to goals, intended student audience, student involvement, content, sequencing, presentation format(s), measures for assessing learning, and teacher involvement, as they reflect theoretical/pedagogical considerations.

(3) study groups in assessing student needs and developing guidelines for selecting appropriate learning materials and outlining adaptations for a specific student group.

(4) demonstrations and workshops devoted to adapting and supplementing selected materials to meet the needs of a specific student group.

(5) workshops in researching, designing, and developing instructional materials for a specific student group.

4. *Convention programming.* In the past few years both national and affiliate TESOL convention programs have included a growing component of demonstrations and workshops devoted to practical aspects of materials development. The enthusiastic reception of these sessions underscores the need for continuation of convention programming devoted to these topics.

More recently, with the introduction of mini-courses at TESOL '77, there is now a format for a longer intensive study session in which non-specialists in an area may benefit from the equivalent of a "short course" taught by a specialist. In future mini-courses, theoretical considerations as well as practical aspects of materials development topics can and should be explored. A variety of annual up-to-date short courses can provide ESL professionals with opportunities for continuing education, (and education for change), relating theory in second language learning and teaching to implications for curriculum design and materials development.

Finally, future convention programming could well include colloquia in which specialists in materials development devote time and discussion to theoretical and pedagogical issues relating to researching, designing, and developing instructional materials for students in a variety of particular circumstances.

There should be an increase in convention sessions devoted to both theoretical and pedagogical aspects of ESL materials development—papers, demonstrations, workshops, mini-courses, and colloquia.

5. Concluding notes

ESL has been undergoing substantial changes and expansions for many years. Changing concepts about the nature of second language learning and learners have influenced language pedagogy and are beginning to be reflected in curriculum design and in the form and function of instructional materials.

There is a serious need for careful profession-wide planning for specialized training in ESL materials development if we are to meet the demands of emerging needs and the challenges of new dimensions in instructional materials. Recent developments in the study of language, language learning, and learner processes should be reflected in new materials but thoughtfully and knowledgeably so that disappointments and disenchantments are minimized so that the new frontiers in materials development are explored carefully not by chance, but by design.

Bilingual Education: Research and its Implications

Merrill Swain
The Ontario Institute for Studies in Education

Bilingual education has different meanings to different people. By bilingual education I mean the use of two languages as mediums of instruction at some stage in a student's educational career. This allows for wide variation with respect to when the languages are used within any grade level or across grade levels. Thus, included as bilingual educational, are programs such as French immersion programs where English-speaking unilingual students are initially instructed in French and do not receive instruction in, or about, their first language until the third or fourth year of schooling, as well as programs for unilingual Chinese students who are taught initially in their mother tongue and introduced to instruction in, or about, English simultaneously or at later grade levels. Bilingual education does not, and should not, exclude the teaching of either language as language per se, but it necessarily involves using both languages as vehicles of instruction.

The sheer volume of words that have been devoted to the topic of bilingual education is, to say the least, overwhelming. One can read about bilingual education from a linguistic perspective, an educational perspective, a philosophical, psychological, sociological, political, historical or a legal perspective (see, for example Spolsky and Cooper, 1977; Centre for Applied Linguistics, 1977). One can find discussions and descriptions of bilingual education in multilingual societies in both developed and developing countries, of bilingual education for immigrant groups, for indigenous populations, for minority language groups, as well as for majority language groups (see, for example, Lord & T'sou, 1976; Spolsky, 1972; Swain and Bruck, 1976). And one can read about research and evaluation studies associated with bilingual education (see, for example Lambert and Tucker, 1972; Cohen, 1976; Mackey, 1972).

However, attempting to come to grips with all the literature, and the contradictory conclusions reached in the various research and evaluation studies, quite simply, boggles the mind! Consider some of the research evidence, for example. On the one hand, evidence has been presented which suggests that bilingual education leads to enhanced cognitive development, high levels of achievement in content learning, high levels of proficiency in second language skills, enrichment of first language skills, increased feelings of self-worth, as well as more positive attitudes towards schooling and towards other ethnic groups. On the other hand, evidence which supports virtually the opposite conclusions has also been presented.

Can these differences be reconciled? I believe they can. Furthermore I think they must be, if our concern is with the provision of educational programs which maximize a child's opportunity for learning and personal growth. What I want to do, then, is to try to provide a basis for understanding the contradictory conclusions associated with bilingual education. In order to do this we will need to consider differences in programs, differences in the children attending the programs, differences in the communities in which the programs operate, and differences in the research strategies employed in the studies themselves.

As one outcome of this exercise, several underlying concepts which in part account for the contradictory results emerge. And it is with a brief discussion of these concepts that I will conclude.

So let me turn now to examine some contradictory evidence. In doing so, it is relevant to note that although studies which have examined the effects of bilingual education have measured a variety of student outcomes, they have all been concerned with at least two questions, and these are the questions we will consider here: First, how do students in a bilingual program perform academically relative to similar students who are not participating in the program? Second, how do the linguistic skills of students in a bilingual education program compare with those who are not participating in the program?

The results that I will refer to come mainly from two different sources: the Canadian data on early total French immersion education and the American data gathered by the American Institutes for Research (Danoff, M. et al. 1978). The Canadian data have been collected by a number of researchers over the last decade or so, each working in their own local area (see Swain, 1976 for a bibliography). The American Institutes for Research (AIR) study, a study commissioned by the US Office of Planning, Budgeting and Evaluation, examined the impact on Title VII students in Spanish-English programs. To do this a large nation-wide representative sample of Title VII projects was selected and compared to similar students in non-Title VII programs. The major part of the study was conducted in the 1975-76 school year and included students from grades 2 to 6.[1] The programs from which the students were selected were in their 4th or 5th year of operation.

With respect to the first question, how do students in a bilingual program perform academically relative to students not participating in the program, studies of French immersion programs in Canada show that students in the French immersion program obtain scores equivalent to their English-educated counterparts on standardized tests of Science and Mathematics, typically scoring at or above their expected grade level in relation to national norms. On the other hand, the AIR study showed that Title VII students either performed similarly or not as well as non-Title VII students, and both groups scored below their expected grade level in relation to national norms in Mathematics. The expectation, of course, was that the Title VII students would do better than non-Title VII students.

---

[1] In the fall of 1976, a sample of classrooms from the grades 2 and 3 cohorts was tested again when the students were in grades 3 and 4.

With respect to the second question, how do the linguistic skills of students in a bilingual program compare with those who are not participating in the program, studies of French immersion students in Canada show that in relation to their first language, there is an initial lag in English language skills on the immersion students relative to their English-educated counterparts. However differences between the groups disappear by no later than grade four. In later grades, immersion students in some cases out-perform their English educated counterparts in various aspects of English language skills, and tend to stand at or above their expected grade level in terms of national norms. It is interesting—and we will return to this point later—that little attempt has been made to measure first language proficiency in the U.S. programs. However, in the AIR evaluation of Title VII programs just mentioned, Spanish reading was measured, and no differences were found between the scores of the students in the Title VII programs and similar students not participating in Title VII programs. The expectation was of course, that students in the bilingual program would do better.

In relation to second language skills, the Canadian French immersion results show that in reading and listening native-like levels are attained, but not in speaking and writing. On a French achievement test standardized on a Quebec francophone population, immersion students score around or below the 50th percentile; whereas, the AIR results showed that on a standardized test of English reading and vocabulary knowledge the Title VII students score around the 20th percentile in grades 2 through 6.

I have taken extremes to illustrate how different the results can be. There are, indeed, other studies which show equally positive or negative results. The issue being raised here, however, is what possible factors might account for the differences in results. How, in fact, can these differences be reconciled?

First, there is the issue of how the data are interpreted with respect to expectations. Second, there are important differences in the strategies employed in carrying out the studies themselves that may account for the different results. And third, there are a variety of background, student and program variables that may also account for the differences. These will be examined in turn.

First, it is important to note that the goals of the two programs are different. In the French immersion program, the goal is to learn a second language without a decrease in native language skills or academic performance. In the Title VII programs, the goal is to learn a second language and increase academic performance. The role of the mother tongue is seen as a bridge so that students are not prevented from learning content material while the second language is being learned. Thus, the results just mentioned are interpreted as positive in the case of the Canadian programs because the goals have been attained. However, in the U.S. the results are interpreted negatively because the expectations have not been met. However it is important to note, as Zappert and Cruz (1977) have done; and I quote them:

> a non-significant effect is not a negative finding with respect to bilingual education. A non-significant effect, that students in bilingual education classes are learning at the same rate as students in monolingual classes, demonstrates the fact that learning in two languages does not interefere with a student's academic and cognitive performance. Students in bilingual classes have the added advantage of learning a second language and culture without impeding their educational progress. Under these circumstances, a non-significant finding can be interpreted as a positive effect of bilingual education.

Let us consider next the nature of the differences between the two sets of studies. Although the AIR study evaluated student performance from grades two to six, it did so by looking at each grade level independently and examined achievement gains over approximately a six month period within each grade level.[2] Although most of us may claim to be optimistic about the effects of education, even the most optimistic amongst us may consider that six months is a rather short period of time to expect significant changes to occur in one group relative to another, especially when one considers that the six-month period is not the only "input" time.[3] That is to say, gains may have already been made by the students during previous years in the program, and these previous gains would not be evident in a comparison of gains over a six month period.

In this respect, the studies on immersion education have been considerably different. The studies have followed students for considerably longer periods of time, in most cases from kindergarten year to their current grade level, thus allowing for the cumulative effect of the program to be determined.

The problem of looking for significant differences in gains over short periods of time in the AIR study is compounded by the fact that very little was known about the previous educational experiences of the students, or their linguistic or attitudinal characteristics at the onset of their experience with bilingual education. We might ask in what ways were the Title VII students different from those who did not enter a bilingual education program? Presumably they were having difficulties with English, yet the results show that by grade 5 they were making gains equivalent to non-Title VII students, and were in fact scoring comparably on standardized tests of English.

Many of the students in the study had been placed into the bilingual program part way through their elementary education. For example, the average number of years previously spent in a bilingual education program for the grade 6 students in the AIR study at the time of testing was approximately three. What was the effect of the switch? Perhaps their parents did not want them in a bilingual education program because they feared their child might not learn English in a bilingual class? Perhaps switching was perceived by the child as a sign of failure in school? Perhaps the Spanish taught and used in class was not their own dialect of Spanish only to reinforce already existing feelings of inferiority?[4]

---

[2] It should be noted that there was a full year between testings of the grades 2 and 3 cohorts mentioned in footnote 1, with approximately similar results.

[3] For the 1975-78 sample, about 75% of the students in grades 2 to 6 had two or more years of bilingual education, and 50% of the students in grades 3 to 6 had three or more years.

And what about the students who were in a bilingual education program in previous years and have since moved to non-Title VII programs because they have been judged to have sufficient knowledge of English to cope with instruction in English? Aside from the fact that their absence would have the effect of lowering the English scores attained by Title VII students, shouldn't they also have been studied as part of the impact of Title VII?[5]

In the Canadian immersion studies, on the other hand, the initial characteristics of the students are known. They are unilingual English-speaking children. They come from middle to upper middle class homes, and their parents have chosen to enroll them in the program. Students do not leave the program when they have sufficient knowledge of the second language, but rather they remain in the program continuing to study in both their first and second languages. There is, however, attrition, and although it has not been studied systematically yet, it is an important question to ask why some students leave the program. In some cases it may be because the program was perceived by their parents or teachers to be too difficult for them, although recent data from Alberta suggests that moving was the main source of attrition. However, just as the "drop-outs" of Title VII programs might perhaps be viewed as a sign of the success of the program, perhaps the "drop-outs" of French immersion programs might be viewed as a sign of its failure.

The point being made here is that without knowing what the initial characteristics of the students were, and the cumulative effects of the program on the students, it is impossible to conclude that the results are either negative or positive. Interpretation of the results as either positive or negative is dependent on knowing what the children were like at the beginning of the program, and observing the effects over time of the program (see also, Cummins, 1979a).

Another difference between the AIR and Canadian studies is that of lump-ing together the results across different program variations. This simply has not been done in Canada. Comparison of results across different program variations have been made on an ad hoc basis (Swain, 1978), and generalizations from similar programs ongoing in different communities have been drawn (Swain, 1979). The AIR study has taken the opposite tact, and averaged results across communities and program formats. Given also the heterogeniety among students in terms of their abilities in each language, their language use patterns, their attitudes toward school and the use of each language from program to program, as well as within programs, the results are basically useless in providing information that would be helpful to program planners. Much more useful would have been to analyze results according to known differences in student and community characteristics and program treatment, examining the individual results to determine if any generalizations could be drawn, even if tentative.

---

[4] D. Reynolds has since pointed out to me "that these speculations are not backed up by any evidence that the students were indeed switching into and out of the Title VII project classrooms for the reasons you suggest. If we are to conjecture about reasons for sixth graders being in the program for fewer than six years, we might suggest that the students came in late because they recently moved into a district with a Title VII program, because they recently arrived in the United States, or because slots in the classroom opened up when the students originally enrolled moved out of the school district. (There is a very high mobility among the student populations in most of the districts with Title VII programs) (personal communication, May 11, 1979). I see no disagreement between the reasons she suggests for why the children were switched, and what I am suggesting—the possible psychological effect of the switch.

[5] It should be noted, however, that "AIR found almost no programs consciously moving students out of the bilingual classrooms once they had learned English. No exit criteria were in effect, and 85% of the projects studied considered themselves native language maintenance programs" (D. Reynolds, personal communication, May 11, 1979).

In effect the AIR study violates the very common sense point that to determine what kinds of programs work best with what kinds of children under what circumstances, one needs to look at the effects of individual programs and community variables as they interact with initial student characteristics over a long enough period that cumulative effects of the program can be observed. Most likely some of the programs were effective at achieving their aims, and some were not. What would be interesting and useful to do is determine the characteristics of the successful and unsuccessful programs.[6]

Finally in terms of research methodology it is worth mentioning a problem that plagues both sets of studies: the instruments used to assess student language abilities and achievement levels. Here, the best that can be said at the moment is that we need to develop much more sophisticated tools for the assessment of linguistic abilities. We must address the issue too, of language for what purpose?—for communication? as an instrument of thought? for self-reflection? for artistic and literary purposes? for what?

What do current language tests measure? Most of the tests used measure vocabulary knowledge, reading and grammatical knowledge. Grammatical knowledge would appear to be a necessary although not sufficient condition for communicative performance, but clearly, tests of grammatical knowledge cannot substitute for tests of communicative performance. Then we must address the issue of what constitutes minimal levels of communicative performance if that is to define limited English (French, Spanish) speaking abilities. And if it is to define the necessary minimal level, then it is essential to consider the relationships between this minimum level of communicative performance and the psycholinguistic concept of threshold level as proposed by Cummins (1978a; 1979a) and which will be discussed shortly.

---

[6] It is the case that in the AIR study, relationships were examined between student achievement and educational processes (e.g. extent of grouping versus individualized instruction, hours devoted to ESL and SSL instruction, use of languages in class), teacher and teacher aide characteristics (e.g. level of education, teaching credentials, years teaching, training in bilingual education), and contextual variables (e.g. proportion of Hispanic students in class, home and neighborhood language use of English) for a subsample of the programs. Although several interesting overall patterns emerge, the problem remains that the analyses lumped together all the programs rather than analyzing individual programs that were successful or unsuccessful and relating these outcomes to the above mentioned variables. And, unfortunately, the analyses used, partialled out the effect of several of the factors which are being suggested here as key factors, e.g. preference for English versus Spanish, and the use of Spanish and English. But indeed, the authors are to be commended for collecting and analyzing such significant data. The major suggestion is that further analyses on a class-by-class basis would be more revealing if the aim is to understand what types of educational treatments are most successful for students of different linguistic and attitudinal characteristics in communities of different linguistic and attitudinal characteristics, given the program goals. In all fairness to the AIR study, however, it must be stated that its aim was not to evaluate any specific approach to bilingual education, but rather to evaluate the overall impact of the Title VII program.

As well as differences in approaches to the evaluation of programs, there are many differences both in the characteristics of the students involved in the Title VII programs and those of the students involved in immersion education. Many of these have already been mentioned in the preceding discussion, but it is important to note the considerable heterogeneity both across and within programs of student characteristics in the U.S. programs, especially as relates to their first and second language abilities at the onset of the program, and the students perception of their participation in the program. The common characteristic among students participating in the Title VII programs as stipulated by the previous legislation was that they were from lower class families. The immersion students, on the other hand, form a much more homogenous group and for the most part, come from middle to upper-middle class homes.

Similarly there are differences in communities—in community attitudes towards the use of the languages involved and towards the program format. Where immersion programs exist, it is usually because the community has fought for it, often against administrator resistance. In effect, the programs have been tailored to meet the community demands. For example, in Ottawa and Montreal where the need for French is perceived as greater than in, for example, Vancouver, more instructional time in French has been demanded. In the U.S. programs, it has often been the case that program formats have also been developed to meet community demands. However, it has also been the case that in some communities the programs have been imposed by legal mandates on communities which are more concerned with having their children learn English than maintain their mother tongue. To them it is anti-intuitional that bilingual education can provide both.

And finally there are program differences. The typical format of a French immersion program is that all instruction is given in French, the second language, until grade 2 or 3. Only then is English Language Arts introduced and taught in English for approximately an hour a day. With each successive year thereafter a larger proportion of the curriculum is taught in English until an approximately equal balance is reached between the time devoted to instruction in the mother tongue and the second language.

The first French immersion program in the public sector began in the Montreal area some 14 years ago. Since then other communities have pushed for an immersion program to be established, due in large part to the reported success of the original program. Each program has begun at the kindergarten level and each year added a successive grade. Thus, the growth both within programs and across Canada, has been slow and steady. Millions of dollars have been spent on the development of curriculum materials, and on evaluations of individual programs by a variety of researchers. The data collected have been fed back to the program developers, leading at times to changes in instructional emphasis and instructional techniques.

The programs in the U.S. are as different as are the students who enter them and the communities they are housed in. The extent to which each language is used, by whom, for what purposes, to teach what, vary from program to program. In addition many programs were implemented before materials were prepared or teachers were trained.

As I have already mentioned, there are differences in the goals of the programs. Title VII is trying to accomplish several things at once: to increase the level of academic achievement, to strengthen self-image and culture identity, and to develop second language skills. In the immersion program, the main aim is to develop second language skills without a decrease in academic achievement or mother tongue development. In the immersion program, no one is concerned with improving the child's self-image or cultural identity; it is not judged to be low in the first place.

Let us consider for a moment a common goal—that of learning a second language. Here I think the Canadian research has an important message to offer and that is, that in spite of the fact that immersion students have "everything going for them"—that is, they come from middle to upper-middle class homes, they have parents who are involved in their education and motivated to have their children learn French, they have native French-speaking teachers and so on—they still have not mastered the spoken and written codes with anything like native proficiency. In spite of their lack of native-like proficiency in speaking and writing, the immersion students have developed sufficient levels of understanding and reading to be able to deal effectively with the material presented to them in the second language. This suggests that limiting the definition of those who need English to profit from instruction in that language to speaking abilities only is perhaps misguided. In light of this, it is a significant step that has been taken by the U.S. Federal legislation to open the Title VII program to those with limited English proficiency in all four skills—understanding, speaking, reading and writing. As I have already mentioned this means that considerable efforts will be necessary to establish what is to be meant by minimal levels of proficiency.

Cummins (1978a; 1979a) has referred to the level which must be attained in a language in order to profit from instruction in that language as the threshold level. Cummins has also suggested that the development of second language competence[7] is a function of the level of the child's first language competence at the time when intensive exposure to the second language begins. This implies that because the level of competence in the second language is dependent on the level attained in the first language, it is important to ensure that the threshold level is also attained in the first language. There is evidence to support this notion from several studies where initial teaching in the mother tongue plus second language classes has led to higher levels of linguistic performance in both the first and second languages relative to those students receiving initial instruction only in the second language. The studies include Finnish immigrants in Sweden (Skutnabb-Kangas and Toukomaa, 1979), francophone minority groups in Manitoba (Hébert et al., 1976) and Alberta (Cummins, 1979b), and Navajo children in Rock Point (Rosier and Farella, 1976). This points to the critical importance of determining the linguistic proficiency of the students in their first language, something which has rarely been done in the U.S. programs because the goals have been oriented to second language learning. But if the potential level to be developed in the second language is dependent on the level attained in the first language, then information about the first language abilities of the children will be important in determining appropriate forms of educational intervention.

---

[7] For Cummins, language competence means "the ability to make effective use of the cognitive functions of language; that is to use language effectively as an instrument of thought and represent cognitive operations by means of language" (Cummins, 1978 footnote 21).

The differences in results, then, can be attributed to a number of factors. They are rather nicely summed up by the notions of subtractive and additive bilingualism as proposed by Wallace Lambert (1975). Subtractive bilingualism refers to situations where the learning of the second language reflects the loss, or "poor" learning of the first language. Additive bilingualism refers to situations where the learning of a second language occurs with no loss to the first language.

For example, in the French immersion program, the children are members of the dominant linguistic and cultural group. Learning the second language does not portend the gradual replacement of the first language and the loss of cultural identity associated with that language. Furthermore the second language being learned is a socially relevant, nationally and internationally recognized language, through which individual economic advantages may accrue to the learner. This situation, that is, where the first language is maintained while a second language is being learned, is an "additive" form of bilingualism. This is in contrast to the situation faced by many immigrant groups, minority groups and indigenous populations who perceive knowledge of the majority school language to be the gateway to social and economic gains, and the home language to be of little consequence except in enabling them to communicate with their friends and relatives, and preserve ethnic identity. The overwhelming use of the dominant language in school and in the wider community often results in a "subtractive" form of bilingualism where the learning of the second language may reflect some degree of loss of the first language and culture.

Combining the ideas of Lambert and Cummins, that is, that the potential for linguistic development in the second language is dependent on the level attained in the first language, and that subtractive bilingualism may indicate lower levels of attainment in the first language—suggests a major implication for education. The implication is that if optimal development of minority language children's academic and linguistic potential is a goal, then the school program must aim to promote an additive form of bilingualism. Attainment of this goal will necessarily involve a home-school language switch at some stage in the educational process, but when, and how much, must be determined in relation to the linguistic and socio-economic characteristics of the learner and of the learning environment. Specifically when the home language is different from the school language and the home language tends to be denigrated by others and selves, and where the children come from socio-economically deprived homes, it would appear appropriate to begin initial instruction in the child's first language with the second language being introduced as a subject of instruction. At a later stage instruction in the second language would then be introduced. On the other hand, where the home language is a majority language valued by the community, and where literacy is encouraged in the home, then the most efficient means of promoting an additive form of bilingualism is to provide initial instruction in the second language (Swain and Cummins, 1979).

These are general statements which now need to be refined by examining specific research hypotheses. Specific hypotheses can be generated which take into account the level of first and second language competence of the learner at the onset of the program, the particular socio-cultural-political characteristics of the environment, and the characteristics of the program itself in determining linguistic, academic and psychological outcomes.

Research has much to offer if we take as our starting point the question: which kinds of children are going to do best in which kind of program in what kinds of communities? Asking the question this way recognizes that there are many possible formats of bilingual education, that communities may differ radically, and that children vary considerably in the characteristics they bring with them to school. Student outcomes must be seen as resulting from community and program variables as they interact with student characteristics.

In summary, the following are the points that I have tried to make:

(1) Different formats of bilingual education are appropriate for different students. The program will interact in important ways with, among other things, the linguistic characteristics the children bring with them to school and community attitudes towards the languages. Research must direct itself to the isolation of other key variables.

(2) Negative results related to bilingual education tend to be associated with situations fostering subtractive forms of bilingualism. The challenge facing educators is how to turn subtractive forms of bilingualism into additive ones. This has been done in some instances, for example, with Finnish immigrants in Sweden, minority francophone groups in Alberta and Manitoba, and indigenous populations in the U.S.

(3) Slow, steady growth of programs, with well-planned curriculum, and strong community and administrative support is bound to lead to more effective results than "shot-gun" implementation.

(4) In trying to interpret the results of research associated with bilingual education, ask yourself the following questions:

(a) What are the characteristics, especially linguistic, academic and attitudinal, of the students before they entered the program?

(b) What are the attitudinal and linguistic characteristics of the community in relation to the program?

(c) What is the nature of the educational treatment especially with regard to the use of the two languages, both as taught and as used as mediums of instruction?

(d) For how long were the students followed? Are cumulative effects of the program being demonstrated?

(e) How is the child experiencing the program, regardless of the program goals?

(f) What was the nature of the tests used? What were they measuring? In what languages were they given, the child's first or second language, the child's dominant or weaker language?

(5) And finally, in relation to the issue of tests, it is obvious that considerable effort is going to be needed to determine what constitutes minimal levels of language proficiency, perhaps most productively pursued by asking what are the threshold levels, that is what does the student need to know linguistically in relation to reading, writing, speaking and understanding in order to be able to profit from instruction in that language?

In conclusion, there is one other point I would like to make. Although I have focussed on outlining some of the reasons why we might find different results in different studies of bilingual education, it is important to note that there is evidence which demonstrates that some bilingual programs have accomplished their aims (see Troike, 1978, for example). It is unlikely that it will ever be possible to determine whether or not the fact that the program is bilingual is the only reason for its success. However, the fact that it has been demonstrated that through bilingual education it is possible to learn a second language and enhance learning at no expense to the mother tongue for minority, immigrant, and indigenous populations is extremely important. It means it can be done! It is possible through the efforts of dedicated teachers and committed communities to change subtractive bilingualism into an additive form of bilingualism. If we are committed to developing to its fullest potential the linguistic, academic and personal growth of the child through education, then bilingual education can provide the means to do so.

Testing Panel

What Makes Test A Four Letter Word?: Viewpoints on Bias in Standardized Testing

Darlene Larson
New York University

Classroom teachers have repeatedly seen evidence of ample ability in their students from linguistically different cultures, yet watched low scores accumulate on their records whenever these students were tested for achievement and ability. They have come to TESOL's Committee on Sociopolitical Concerns year after year to beg that we "do something" about this inequity. The time was long overdue, and with the cooperation of the Second Vice President Carlos Yorio, the Committee on Sociopolitical Concerns was able to turn the attention of TESOL '79 participants to the issue of bias in standardized testing.

Opportunities for bias in testing are unending. Gilberto Garcia, then of Georgetown University, called to our attention at TESOL '78 the many cases of bias in the routines and procedures at the testing site, and in the scoring and the way the results were used. Both in the judgment of individuals' records, and in the guidelines and policies of federal and state education agencies, we must be alert to the implications of every regulation.

Panelists at TESOL '79 whose remarks follow are: Dr. Ernest M. Bernal, Jr. of the Educational Testing Service in Austin, Texas, Dr. John W. Oller, Jr. of the University of New Mexico, Dr. Paul Rosier of the Page, Arizona Public Schools and Dr. Roger W. Shuy of Georgetown University. The remarks of our fifth panelist, Dr. Edward DeAvila of Stanford University were not available for inclusion. Dr. DeAvila's views can be found, however, in several articles such as, "Piagetian Alternative to IQ: Mexican-American Study" by Edward DeAvila and Barbara E. Havassy in *Issues in the Classification of Exceptional Children*, Nicholas Hobbs, Editor, Jossey-Bass Publishers, San Francisco, 1975.

We in TESOL are eager to see more attention paid to finding new kinds of assessment instruments and assessment policies which will maximize the opportunities for success for every student, pre-kindergarten through adult, for native speakers, non-native speakers, and for those who function in standard English as a second dialect.

I wish to acknowledge my debt to an earlier article by John Upshur, "Test is a Four Letter Word," for the genesis of my title.

Testing Panel

What Makes Language Test an Expletive: Tests, Test Selection, and Test Use

Ernest M. Bernal, Jr.

Educational Testing Service

1. Tests

The expanding fields of bilingual education and ESL have created a considerable demand for language assessment instruments for many age levels and different educational settings. Many tests, of course, employ the more traditional norm-referenced scales and discrete-point, multiple-choice formats to measure a variety of related verbal skills. As a member of a state language proficiency instrument review committee, however, I have noted that many new tests carry the names of linguists as authors, that language-eliciting procedures more characteristic of linguistic research methods are being adapted to widespread language testing, and that greater attention is being given to criterion-referenced levels of competency using integrative or semi-integrative formats. Fortunately, too, there are tests which seem to incorporate features from both "camps," as it were (e.g., multiple-choice cloze tests).

Yet it is this very demand which has contributed to certain abuses. One might say that a bandwagon effect obtains, that certain individuals and test publishers are rushing the completion and dissemination of language tests in order to capture a profitable percentage of the market, even though many of these tests lack the qualities necessary to recommend them (See, for example, Silverman *et al.*, 1976).

For example, at least two "tests" do not require children to respond or even to be observed. They are essentially demographic questionnaires which can be filled out using school records or through an interview with an adult family member. What's worse, no effort has been made to validate these instruments against behavioral criteria, viz, how a child actually uses language. One of these instruments appears to have grown out of a family linguistic interaction study.

Other tests either lack a theoretical rationale or are based on as-yet unverified hypotheses about language acquisition or performance. One technique assumes that the ability to name the antonyms of individual lexical items in two languages is a good measure of bilingual proficiency, another uses the ability to repeat phrases without errors in pronunciation as adequate to measure language dominance. And the list goes on. There are tests which hardly require a person to speak at all but only to indicate comprehension, yet which are interpreted to indicate oral performance.

A third category of tests measure only marginally useful constructs for educational purposes or confuse the constructs they purport to measure. Tests which measure language dominance without yielding an estimate of comparative language proficiency are actually misleading to educational decision makers. Many practitioners do not realize that one can be dominant in a marked language while still competent in the unmarked tongue. In the same vein, procedures used to determine a person's "preferred" language are not necessarily measures of language proficiency.

Another popular practice is to develop a test for local consumption and then to market it nationally, despite its obvious dialectal biases. This cleverly shifts the burden of validity entirely to the out-of-region purchaser, who very often has to "adapt" the test anyway, thus further confusing the issue by admitting non-standardized testing procedures.

Finally, many of these problems with tests could be avoided if otherwise naive test makers would follow or devise some rigorous system of test development. The responsible test developers-publishers have articulated adequate systems of test development and generally make them available for the asking. (See, for example, ETS, 1979.) These guidelines can be characterized by research and development (R&D) procedures, in which the ultimate touchstone of a test is the empirical analysis of how it fares in the field.

Certainly one of the principal reasons why test is a four-letter word, why so many tests evidence such strong biases against minorities—even to the point, in some instances, of not even measuring the same construct (Anastasi & Córdova, 1953)—is precisely this lack of development. The target population(s) are inadequately specified, insufficient trials are conducted, revisions are not made on the basis of empirical analyses of answers, different formats (e.g., different directions, pictures, answer sheets) are not tried out, validity is left entirely to someone's "expert" judgment about what constitutes a good language sample, and levels of competence are set without reference to the abilities of native speakers. It happens, for example, that some linguist's research instrument appears quite suddenly in printed form for some more general educational use. In short, too many language tests are not developed but merely produced.

2. Test Selection

Yet another reason why test is a four letter word has to do with the way tests are selected (or kept in use). Too often test selection is a decision based upon a single factor, such as cost, administrability, scorability, attractiveness, popularity (i.e., "Everyone uses this test"), or political favoritism (i.e., "George, the salesman, is a good ol' boy" or "June, the author, is a linguist, so this test must be better than the one devised by all those nasty psychometricians.")

What's worse, teaching and supervisory personnel are rarely consulted or trained in these matters, thereby heightening the likelihood of hostility toward or mistrust of a particular test. Needless to say, facile test selection frequently results in the use of inappropriate instruments for the populations and school programs of interest.

Some test manuals also report inappropriate statistics or incomplete statistics, and one is left to wonder about the adequacy of the test. Inter-rater reliabilities, for example, are important to report when a test requires an examiner to transcribe what an examinee says or to make judgments about the accuracy of his/her utterances. Estimates of the internal consistency of a test, while appropriate for language proficiency tests or techniques which are unspeeded or at least not highly speeded, must nevertheless use a special technique whenever the individual "items" are not behaviorally independent of each other (American Psychological Association, 1974) as in the case of cloze tests. Also, it appears to me that the test-retest method of estimating reliability, which is not particularly useful to most traditional psychometric applications, merits further attention in language proficiency measures where an examinee must formulate a verbal response.

3. Test Use: Misuses and Abuses

Once a test is adopted many institutions find it "inconvenient" to remove it from use. In some cases, institutions and researchers continue to use obsolete tests which even the test publishers no longer market. The point to remember here is that unsecure tests, like dairy products at the super market, have a certain "shelf life": Their validities "spoil" as their content becomes known and outmoded. Tests must be either revised periodically or relegated to the shredder. The *MLA Foreign Language Proficiency Test for Teachers and Advanced Students*, as an illustration, is no longer published by ETS, yet one can still find yellowed copies of it being used to screen bilingual teacher candidates!

Other sources of misuse stem from the mindless, routine testing which actually benefits no one. Incredible amounts of money, for example, are spent in perfunctory evaluations and assessments. One of the "necessary evils" which teachers and administrators have unfortunately learned to live with is the testing associated with evaluations intended merely to keep some governmental agency pacified (i.e., by meeting its requirement to conduct and evaluation). The resultant mountain of test data is never utilized to help individual students or to review the program's functioning in detail. Test results are very often not reported at all to instructional and supervisory personnel, or are reported in an untimely manner. I hasten to point out that even tests devised by linguists are subject to these kinds of misuses.

Tests are abused when they are applied to persons for whom they are not appropriate, when they are used without verification in situations for which they were not intended, or when their results are overgeneralized. Certain language proficiency testing formats may not be generally efficient for all American or foreign populations; a test of English language proficiency may not be a valid tool for relegating minority students to non-academic curricular tracks; nor can a person who doesn't speak much English be considered "culturally deprived." The roots of these abuses can generally be traced back to researchers' and practitioners' misunderstandings of what tests are and their misinterpretation of the measurements which tests yield. (See American Personnel and Guidance Association, 1978.)

4. Concluding Remarks

I have departed somewhat from my originally prepared remarks not to condescend to present elementary points about testing but to share some of the fundamental and hard won lessons of my profession with you. During this conference I have heard a number of scholars in the several subdisciplines and specializations of linguistics and language teaching make statements about tests and testing which are basically in error. (A test which reports a reliability coefficient of .90 does not mean that the test score is "wrong" ten out of every 100 times!) What's perhaps more telling is that their statements seemed to go unchallenged except by those of us in attendance whose training or experience has steeped us in test development.

The other night a few thousand of us heard a fascinating set of characterizations (or is it *caricatures?*): Psychometricians are supposed to be preoccupied with reliability, while linguists are concerned more centrally with the issue of validity. Such a heuristic is neither accurate nor constructive. I believe it is devisive, and that it should be examined.

In the first place, there seem to be increasing numbers of language and testing specialists who are crossing over to learn from one another. Secondly, test reliability is propadeutic to test validity, because test reliability refers to the degree of confidence that one can place in an obtained score or classification. Tests which evidence low reliability leave one in serious doubt as to where to place an examinee on a scale of language comprehension, for example.

Finally, validity is not all of a piece, as it were. There are many kinds of test validity, and a test which doesn't meet a linguist's criteria for content validity may nevertheless have excellent predictive validity and utility for certain consequential decisions, such as educational placement in a particular language program.

And here, perhaps, we can join the issue and learn from one another. One of the true characteristics of psychometricians is their empirical bias; they have a real "show me" attitude towards any test or technique, and it is my devout wish that you will respect this dear if sometimes chaffing propensity, for it is founded on costly experience. If linguists indeed can produce better language tests than non-linguist psychometricians, then it is because they have devised demonstrably superior methods of obtaining language data and ensuring its more accurate interpretation and application to issues and problems which are worthy of measurement, the professional concerns of the TESOL membership.

So I now propose renewed efforts for us all to collaborate, not because psychometricians are not producing valid language tests, but because together linguists and psychometricians might increase the validity and utility of language measures while satisfying the requirements of our respective disciplines.

Testing Panel

Intelligence, Achievement, Competency, or Language?

John W. Oller, Jr.

University of New Mexico

A sagacious pre-biblical Pharoah gave the following advice to his successor son sometime in the third millenium before the first coming of Christ: "Be a craftman in speech [so that] thou mayest be strong . . . the tongue is a sword . . . and speech is more valorous than any fighting . . ." (Lafay, Stanfield, and Glanzman, 1978). These words were inscribed on a cuneiform tablet recently found among the ruins of the great empire of Ebla. Sometime later a Chaldean named Abraham settled in that same land, and some centuries later still one of his children wrote "The Gensis record: 'In the Beginning God created. . . .' In the beginning was the Word . . ." Much later still, but less than a century after the crucifixion, John paraphrased and expanded on the Genesis account writing and wrote again, "In the beginning was the Word, and the Word was with God, and the Word was God" (John 1:1, New American Standard). And in verse 14 of the same text he wrote, "And the Word became flesh and dwelt among us . . ." (John 1:14, New American Standard).

More recently, just before his death, the Russian psychologist Lev Vygotsky (1934) avowed that Faust was more correct when he said, "In the beginning was the deed." Vygotsky contended that "action was there first" and that the word "is the end of development, crowning the deed" (p. 153). Of course, neither Moses nor John were writing about child development, but rather about the creative intelligence responsible for the existence of the universe. In either case, there is a fundamental problem in trying to put actions ahead of the organizational power of the word. Is it possible to conceive of meaningful action without a plan? There is a serious question whether any order can exist biologically, physically, or otherwise apart from something very much like the propositional forms of language. Consider for instance the much celebrated words and phrases of the genetic code, or the conversational exchanges between the cells during embryonic development. Apparently geneticists do not doubt that such conversations take place. The debate only concerns whether the conversations are long and interesting or short and boring (Wolpert, 1978). Or consider the evidence from the research of T. G. R. Bower showing that an infant apparently knows even at the instant that it emerges from the matrix that a visible finger is tangible (Bower, 1974). Even an act as simple as reach-ing and grasping, as was pointed out long ago by the famous neurologist, Karl Lashley, involves the sort of propositional planning (agent, instrument, action, object) that is characteristic of the sequential and hierarchic ordering of words, phrases, sentences, paragraphs and discourse in natural languages.

An interesting question to ponder is whether the order that is observed in the physical universe, the genetic code, or the knowledge of an infant at birth can be characterized in any form other than the propositional structure of natural language or some logical equivalent of language. We might well ask whether it is possible for there to be order or a plan apart from the organizational force of a kind of propositional logic. Of course, it is possible for an organism to behave in an orderly manner without being able to explain its actions in words, but even the simplest responses of a unicellular organism such as an amoeba—say, in moving away from heat—have a kind of propositional character. The organism is *programmed* to respond to certain states in specified ways. Whether the program is acquired or innate is not the issue. The nature of the program is the issue. Could it be construed as anything other than a kind of propositional calculus or logic?

There are many avenues of approach to the fundamental questions we are considering here. Perhaps the most powerful method of all is the exercise of intelligence in theory building. Empirical methods can also be applied at least indirectly by means of specific hypotheses derived from the general theoretical considerations. One of the most appropriate empirical methods of addressing these matters is to examine psychological tests. If the deep logic of language is in fact fundamental to all sorts of actions and systems of order, it should certainly be the case that tests requiring complex problem solving or systematic and sequential cognition should be substantially related to tests of the ability to process language. Putting the matter more precisely, it should be possible to demonstrate a substantial communality across tests aimed at purely verbal processing, such as filling in blanks in a cloze text where every *n*th word has been deleted from a text, and tests aimed at non-verbal processing such as solving visual puzzles or seeing similarities or complex relations in visual patterns, and so forth. In fact, we should find a substantial overlap in variance (in the statistical sense) in tests of widely different sorts such as taking dictation or doing arithmetic problems.

A few years ago when some of these questions were raised seriously for the first time, two types of reactions prevailed. Some of the theorists (so-called "experts") said that language tests could not possibly predict very much of the variance in non-verbal tasks and others said so what if they could. Hardly any of the educational and psychological specialists would admit the seriousness of the question. Some were willing to concede the point on the presumption that one has to understand and produce intelligible language in order to get along in a complex modern society, but this is known by everyone and doubted only by fools. So what!!! Other specialists followed an equally superficial logic considering only the outward appearance of tests and reasoning that no matter what any theory might suggest to the contrary, doing arithmetic is not the same as filling in blanks on a cloze test and that's that!!! A few specialists even claimed that the importance of language ability to tests of all sorts was so well established and so well researched that no further work was needed. They harked back to the work of the giants in educational measurement and their acquiescence concerning the importance of "verbal comprehension" to all kinds of tests.

All of the foregoing arguments, however, were based on an exceedingly impoverished understanding of the nature of language proficiency. This is not to say that what is known today is the final word by any means, but it is certainly to be expected that recent advances in linguistics, psychometrics, and language testing research have increased our knowledge of the nature of language ability. Interestingly, many of the surprising discoveries of recent years have come from the field of second language testing research. In fact, findings in this field, which defied the predictions of prevailing theories, set the stage for the present controversy over the validity of educational and psychological measures of all sorts.

Classical theories of tests and measurement led the initial theorists in foreign and second language testing to expect certain outcomes which have never materialized in any of the empirical studies. Naive structural linguistics and contrastive analysis also contributed to some of the speculations that were bandied about in the literature as dependable truths. For instance, the popular method of testing foreign language skill known as dictation was rejected by Lado (1957) on the grounds that it did not segment the various elements of structure and aspects of skills in the manner required by what was later called "discrete point" theory by Carroll (1961). According to Lado, and others, it was essential to divide language up into its component parts and test these components one by one. For instance, vocabulary, syntax, and phonology all had to be tested separately as did productive and receptive repertoires, as well as literacy and what has come to be called "oracy". In fact, each element of many different inventories subdivided and subclassified in various ways each had to be tested separately. For example, at least in theory, it was necessary to assess receptive vocabulary in literacy and productive vocabulary in literacy; receptive vocabulary in oracy and productive vocabulary in oracy; etc., etc., etc.

This kind of thinking did not originate in the field of second language testing. In fact, it went back to the heyday of intelligence testing and the writings of people like L. L. Thurstone and T. G. Thurstone (1941) who insisted that complex mental abilities could be decomposed into multiple primary abilities. They argued for the description of mental ability in terms of a "profile" which would display an individual's disparate potentials on a variety of primary abilities measured by separate subtests in a battery of intelligence tests. They wrote,

> there is nothing wrong about using a mental age or an intelligence quotient if it is understood as an average of several tests. The error that is frequently made is that

> the intelligence quotient is sanctified by the assumption that it measures some basic functional unity, when it is known to be nothing more than a composite of many functional unities (p. 8).

L. L. Thurstone was particularly concerned to dispute the argument of Spearman (1904, 1927, and 1950) who had produced strong empirical and statistical evidence for a general intellectual ability that seemed to be rather unitary.

The point here is not to defend Spearman's theory of general intelligence, but to point out the connection between the kind of analytical thinking that came to pervade the whole field of psychometrics and the discrete point approaches to language testing. We will return to the general intelligence factor below and propose that it is identical with a general factor of language proficiency. In the meantime, it remains to consider some of the empirical findings of second language testing research that contradict the claims of discrete point theories. The treatment here is necessarily sketchy and incomplete as other works have dealt with this matter in much greater detail (Oller, 1976b, 1979b).

Until about ten years ago, little empirical research on language testing had been done. Rebecca Valette had shown that dictation was about as good a measure of overall proficiency in French as a more elaborate objective examination as early as 1964, but the application of correlation and more sophisticated statistical techniques was almost unknown to the field of language teaching and testing until the late 1960's and early 1970's. In 1969 it was discovered that the dictation portion of the UCLA *English as a Second Language Placement Examination* (Form 1) correlated more strongly with each of the other subscores than any of them did with each other. This finding was rather surprising as it contradicted the expectations of the popular theorists. It showed that something about the dictation procedure was capable of generating more meaningful variance across individuals than any of the other tests (Oller, 1970). The puzzling question was why should dictation correlate so strongly with tests that didn't even require auditory processing of speech signals?

The unexcepted strength of the correlations observed between dictation and other testing methods was subsequently borne out in repeated studies (Johansson, 1973, Oller and Streiff, 1975, Whitaker, 1976, and others). Moreover, several other testing techniques were later discovered to have similar robust properties. For instance, the cloze procedure which had been widely studied with reference to reading ability in first language research proved to be an exceedingly robust tool for testing language profficiency in general (Oller and Conrad, 1971, Oller, 1972, Stubbs and Tucker, 1974, Hinofotis, 1976, and many others). Swain, Dumas, and Naiman (1974) obtained similar results with elicited imitation (simple repetition) as a testing tool. In 1976-1977 at Southern Illinois University's Center for English as a Second Language, a number of other testing techniques were added to the list (see Scholz, Hendricks, et al, 1979). Controlled essay and oral interview proved to be about as effective as dictation, cloze, and elicited imitation, and other techniques are still being added. For instance, Robert Scott, Susan Rubin, Mary Galvan, and Gustavo

Gonzalez at Resource Development Institute in Austin have produced evidence showing that a command test in the tradition of J. J. Asher's approach to language instruction is also an extremely effective testing procedure.

Research at RDI in Austin shows that a wide variety of testing techniques such as those mentioned in the preceding paragraph generate reliability coefficients in the high nineties and validity coefficients in the high eighties and low nineties. How is it that such global testing procedures can be so robust and effective? Furthermore, how is it that such radically diverse techniques as oral interview, written cloze, dictation, essay, elicited imitation, imperatives, and the like correlate so strongly with each other? Could they all be tapping the same ability?

The popular discrete point thinking of the 1950's and 1960's could not encountenance the results of the research. From the vantage point of the discrete point theories, there simply is no explanation for the strong correlations that have been found consistently across radically different methods of testing. Moreover, discrete point theorizing resulted in many predictions that turned out to be incorrect. For instance, if a vocabulary test in a written mode is really measuring something other than what is measured by a grammar test, how is it that both may correlate more strongly with a dictation or a cloze test than they do with alternate vocabulary and grammar tests? Discrete point theory can offer no explanation for the fact that tests aimed at different components are often apt to correlate as strongly or more so than tests aimed at the same components of language proficiency. More problematic still is the fact that a test that is not even aimed at any single component (such as a cloze test or a dictation) may measure each component more validly than do the tests that are designed to do so. For instance, a phonology test or a pronunciation score may correlate more strongly with a dictation, say, than with another phonology or pronunciation score. (See research cited above and references there, also see Oller, 1979b).

As the results came in from year to year, it soon became clear that there was something radically amiss about the analytical views that discrete point language testers had adopted from intelligence testing literature. Why should tests that appear very different at the surface produce results that are quite similar while tests that appear very similar are apt to yield less similar results? How is it that the analyses of language skill into so many categories was leading to unexpected outcome? An answer to these puzzles was not long in coming.

In fact, long before the results of the language testing research began to disturb the discrete point theorists, there were sound theoretical reasons to expect those results and to reject the thinking of the discrete point approaches. Who would propose that you can assess the efficiency of a machine by taking it apart and testing the parts without reference to the function for which the machine was intended? Surely if such an approach is not applicable to pure mechanisms it is even less applicable to something as complex as a human language user. Imagine trying to test the efficiency of a racing car by rolling all four wheels down the track at top speed without the rest of the car! Is this any more absurd than testing phonemic contrasts and claiming to test language ability? Or testing surface morphemes and claiming to measure language ability? It would not even make sense to add the scores on a great diversity of discrete point tests and call the total a measure of language proficiency. Adding a bunch of invalid scores does not produce a valid score. If discrete point tests happen to produce some meaningful variance and the composite of several of them turns out to be fairly reliable, it will have to be purely accidental because the theory on which the various subscores are based is quite simply incorrect.

Why should tests that appear very different at the surface produce very similar results? If we consider what is to be learned from a pragmatic theory of language use it is obvious that surface form is a subordinate and in many case inconsequential aspect of language in use. For instance, suppose you receive news that the bank will make the loan you are applying for to purchase a house. Does it matter whether the message is delivered by phone, or mail, or face to face? Is the number of plural morphemes used in the message in anyway significant? Suppose the bank officer, whom you know, merely calls and says, "You're go on the house," wouldn't this be just as effective as a lengthy letter? Does it matter if a passive or active sentence structure is used? The obvious fact is that the surface form of a message is largely unimportant. It is the meaning or pragmatic import of the message that matters.

Consider the possibility that the caller might repeat the message in exactly the same words ten times—or perhaps even repeat the call ten times. Would each of these repeated surface forms have the same consequence? Absurd! The point is that the surface form is consequential only insofar as it affects the message or helps or hinders the communication. Even identical surface forms do not have identical consequences from a pragmatic point of view, and totally distinct message forms in completely different modalities may in fact have (for all practical purposes) identical pragmatic consequences.

Therefore, we should not be surprised by the finding that there is a powerful relationship across language processing in different modalities. In a very important sense, the problem of communication is the same whether the message is written or spoken and regardless whether the task is to send it or interpret it. Even when the purpose of language use is not so much to convey information as it is to enjoy phatic communion or private revelry, the focus of attention is still away from phonemes and morphemes and gramatical structures per se. As the earliest writers on pragmatics insisted many years ago, the totality of a language experience is greater than the sum of the distinguishable parts. It is not in fact the parts per se that have importance but rather their arrangement in relation to each other and to experience—their temporal sequence and their inferential connections to states of affairs. The essence of meaningful language use transcends the surface form of the words in much the way that the human body is more significant than the clothing we happen to put around it, or in the way that the personhood of the human being transcends the physical stuff of which the body is made. Wouldn't it be just a little pretentious for an anatomist to claim he could fully describe the nature of the human organism in terms of measurements of the various parts of the bodies he had studied in his laboratory? Is it any less pretentious for linguists and educators to claim they can measure language proficiency in terms of a superficial analysis of utterance forms and sentence patterns?

Apparently, the reason that pragmatic testing procedures work as well as they do is because they bear some resemblance to the sorts of applications that people make of language in normal experience. In fact, such tests meet two naturalness criteria: first, they require the processing of verbal material under normal temporal constraints; and second, they require the pragmatic mapping of that verbal material onto experience. Put more simply, they require the time constrained processing of meaning.

There are reasons, therefore, to expect certain testing procedures to work considerably better than others. In particular, there is no reason to suppose that a test aimed a phoneme discrimination will be a particularly good measure of anything of consequence—certainly not the ability to use the phonemes in question in normal communication. On the other hand, there is every reason to expect testing techniques such as cloze procedure, dictation, elicited imitation, essay, oral interview, and other methods which meet the pragmatic naturalness criteria to prove to be highly useful measurement approaches. Further, we should expect the latter techniques to be strongly correlated with each other. Just how strongly is perhaps the most important outstanding research question in current studies.

In 1976 it was suggested that perhaps the general factor of intelligence proposed by Spearman (1904) to explain the overlap in variance on many complex problem solving tasks was really a language proficiency factor. At that time considerable evidence from second language testing research was offered showing that a single language proficiency factor could account for the vast majority of the variance in a number of diverse testing procedures (Oller, 1976a). Since the testing procedures for measuring language proficiency were considerably better researched at that time than they had been in the early 1900's when Spearman proposed his theory of general intelligence, it was supposed that perhaps what Spearman thought to be a general cognitive capacity was more accurately characterizable as a global factor of language proficiency —an acquired kind of knowledge or skill. This is not to minimize the importance of the factor nor to propose that it is not an essential ingredient in learning, but it is to say that whatever the tests measure may be much more closely related to background and experience than it is to genetic inheritance. If this hunch were correct, it would remove any vestige of empirical support for the Jensen-Herrnstein hypothesis about the heritability of IQ. No one can reasonably claim that proficiency in a particular language is inherited. The capacity to acquire language surely is inherited, but this is not what we are measuring when we assess acquired language proficiency—at least we are not measuring the inherited capacity directly, only its results. Moreover, we know that language proficiency is highly malleable and is largely determined by experience.

The first empirical study with monolingual native speakers was carried out in 1976-1977 in Saint Louis by Thomas Stump. He asked how much of the variance in the Lorge-Thorndike verbal and non-verbal IQ scores, and in the several subtests on the Iowa Tests of Basic Skills would be attributable to a language factor measured equally well by cloze and dictation. He tested about 100 children at grades 4 and 7 who were native speakers of English. His results showed that 54% of the total variance in all of the tests at the fourth grade level was attributable to a language factor and 62% at the seventh grade level. Interestingly, 72% of the variance in the Lorge-Thorndike non-verbal score was attributable to the language factor for seventh graders. The strength of correlation across tests was particularly remarkable due to the fact that the cloze and dictation testing was done a little more than six months after the IQ and achievement testing.

The second study was carried out by Virginia Streiff in response to Stump's lead. In 1977 she analyzed data from Hopi-English bilingual children employing the subscores of the California Achievement Tests and both an oral and a written cloze test. Her results with 47 children from grades one through six showed that a single factor accounted for no less than 79% of the total variance in all of the tests. The CAT arithmetic computation score (not the word problem score) correlated at .92 with the single factor extracted. That is 85% of the variance in that subscore was attributable to a language factor. As in the case of Stump (1978), Streiff's language tests were administered some months after the CAT. However, in 1978-1979 she was able to replicate her study with nearly simultaneous testings and found that a single language factor accounted for 85% of the total variance in the tests. The variable that correlated strongest with that factor was the written cloze test (no oral test was included in the replication). For the 60 children tested, ranging this time from second through sixth grade, 90% of the variance in the general factor was attributable to the cloze score and no less than 77% to any other score. The arithmetic subtests aimed at pure computation and word problems respectively correlated at .90 and .94 with the language factor. That is, 81% and 88% of their respective variances were attributable to the language factor. (See Streiff, 1978, and 1979.)

Initially, the possibility of a global factor of language proficiency was proposed with respect to second language data. The findings of Stump (1978) extended the argument to first language data and those of Streiff (1978) to data from bilinguals. At about the same time that the Stump project was being carried out, Flahive (1977) was conducting one of the first studies attempting to separate non-verbal intelligence from language proficiency in non-native speakers of English. In spite of the fact that he continued to argue in favor of a distinction between non-verbal IQ and reading ability, his data showed that in fact the distinction is highly suspect. The non-verbal test he used was one of the measures referred to by Jensen (1969) as a nearly "pure" measure of "g"—or general intelligence—namely, Raven's Progressive Matrices. Flahive also used the McGraw-Hill Basic Skill System Reading Test, the Test of English as a Foreign Language, the Perkins-Yorio paraphrase recognition test, and a cloze test. For 20 advanced non-native speakers of English, correlations across tests ranged from .59 to .84. Interestingly, the strongest correlations were between the Raven test and the McGraw Hill (.84) and the TOEFL with the Perkins-Yorio test (.84). Why should tests aimed at the same construct correlate only about as strongly as tests aimed at totally different constructs—e.g., reading proficiency in a second language, and non-verbal IQ? It is especially surprising that non-verbal IQ might be quite indistinct from proficiency in a *second* language. Nevertheless, it is known that proficiency in the first language is strongly correlated with proficiency in a second (Bezanson and Hawkes, 1976, Swain, Lapkin, and Barik, 1976, Johansson, 1973, and others).

Does Flahive's study provide evidence for a single general factor of proficiency including the non-verbal IQ test? Empirical study shows the answer to be a resounding yes. Table 1 presents the results of a principal components analysis (with unities on the diagonal) of the original correlation matrix from Flahive's data. The loadings on the first principal component account for so much of the total variance that it is highly doubtful that any reliable variance would be left in any of the tests if this general component were extracted. The proportion of total variance accounted for by the single component is 74.9%. It is important to remember that this analysis is completely unique and determinate (Nunnally, 1967: 317). If we were testing the hypothesis that a single general factor accounts for the vast majority of the variance in all of the tests, we would have to accept that hypothesis.

TABLE 1

A Principal Components Analysis of the Correlations Reported by Flahive (1977) between Various Tests Administered to a Group of Advanced Second Language Learners
N = 20

Tests	Loadings on (or Correlations with) a General Factor, g	Proportion of Variance in Each Test Explained by g
Raven's Progressive Matrices	.864	.747
Test of English as a Foreign Language	.877	.769
McGraw Hill Reading Test	.866	.751
Perkins-Yorio Test	.883	.779
Cloze Test	.836	.699

While Flahive employed non-native speakers of English, his findings in concert with previous findings would lead us to suppose that what has formerly been general intelligence might actually be a language proficiency factor. At the Boston TESOL meeting in 1979, Chesarek and Oller presented data from a bilingual population of English-Crow speaking children from Montana. Tests included not only the Raven Progressive Matrices, but also Cattell's Culture

Fair Intelligence Tests, Scale 2, Forms A and B. These tests are all aimed at non-verbal intelligence. Also included were the Peabody Individual Achievement Tests (Mathematics, Reading Recognition, Reading Comprehension, Spelling, and General Information), the Peabody Picture Vocabulary Test, several subscores from the Illinois Test of Psycholinguistic Abilities (Auditory Reception, Auditory-Vocal Association, Visual-Motor Association, Verbal Expression, and Grammatic Closure), and a syntax test developed by Ursula Bellugi. Results from first, second, and third grade children revealed large and stable loadings on the first principal component extracted by the principal axes method. Moreover, on the average no reliable variance was left in any of the tests if the general factor was taken out. The mean specificity (i.e., the proportion of reliable variance unique to each test) was —1.1%, while the average proportion of variance in each test that was attributable to a global language factor was no less than 54% of the total. Although the numbers of subjects studied in each sample were not large—they ranged from 22 to 78 in the three samples—the results do not support the common view that non-verbal skills can clearly be distinguished from language ability. Specificity estimates for the non-verbal IQ tests were highly unstable ranging from a low of —.16 to a high of .14.

In sum, data from the bilingual population studied simply does not support the common claim that multiple primary factors (see Thurstone and Thurstone, 1941) can be distinguished. In particular, the argument for a clear separation of verbal and non-verbal abilities seems to be strongly refuted. But what about the case with native speakers? Isn't it possible that a bilingual population might produce results radically different from those in a monolingual population? To answer these questions in part, data from the Illinois Test of Psycholinguistic Abilities were examined. The Manual to the experimental edition reported results of a principal components analysis of the nine subtests in the original battery based on a sample of 700 children ranging from 2½ to 9 years of age. All of them were presumably native speakers of English. A least four of the subtests could be construed as requiring "non-verbal" processing of various sorts—Visual Decoding, Motor Encoding, Visual-Motor Sequencing, and Visual-Motor Association. These tasks required classifying pictured objects (using a simple matching procedure), showing how certain objects are used (e.g., a gun, or a violin), placing objects in a previously displayed arrangement, and indicating which objects go together, respectively. Specificity estimates for these tests ranged from a low of —.24 to a high of .09. Again, the specificities were highly variable and unstable whereas the loadings on the first principal component were large and stable accounting for 80% of the total variance in all nine subscores. The mean specificity estimate, however, was a —.1%.

The data so clearly support the hypothesis of a single global factor of language proficiency that there cannot remain much question about the existence of such a factor. The question that does remain is whether there is anything else left if this factor is extracted from the reliable variance in the tests.

It also remains to extend the findings from bilingual and monolingual children to bilingual and monolingual adults.

At the 1979 Boston TESOL meeting, Paul Streiff presented findings from Navajo youngsters up to the ninth grade. He investigated the various subscores on the Comprehensive Tests of Basic Skills along with written cloze tests. Approximately 50% of the variance in all of the tests was attributable to a single language proficiency factor across each of the three samples examined. Interestingly, there was a tendency for the amount of variance attributable to a general language factor to increase as children progressed through the grades. This result was consistent with the findings of Stump (1978). For a more complete discussion, see P. Streiff (forthcoming).

While there is still a gap in our knowledge at the secondary and post-secondary level for bilingual populations, data on high school age young adults who are monolingual speakers of English is available for the Adult Performance Level Survey (Oller, 1979a). The APL is among the popular new breed of tests aimed at so-called "real life competency". In fact, this instrument is the prototype of competency tests in much the same way that the Stanford-Binet was the prototype for IQ tests. It was pretested on 2,853 subjects from 36 states. The subtests in the APL battery are aimed at two categories of competency: content areas (including community resources, occupational knowledge, consumer economics, health, and government and law) and skills (identification of facts and terms, reading, writing, computation, and problem solving). The correlation matrix for the original test population (reported in the Manual) was examined by several statistical methods. All of them revealed a large variance overlap across subtests. A principal factoring method (distinct from principal components analysis in regard to the handling of diagonal elements) with estimates of reliable variance on the diagonal of the correlation matrix showed 100% of the reliable variance (within rounding error) to be attributable to a single general factor.

How can we be sure that the general factor identified in all of the preceding studies (and others not cited) is in fact a global factor of language proficiency? Could it not be an intelligence factor along the lines of classical theory? There are actually several sorts of evidence and argument that can be brought to bear on these questions. Among them are simple examination of the correlations with the general factor, content analysis of the tests, and theoretical reasoning. All of these sources of argument suggest that the general factor we have been talking about in all of the foregoing studies must be a language proficiency factor. For instance, the tests that correlate most strongly with the general factor (or first principal component) in the statistical analyses are consistently language processing tasks. For example, the test producing the strongest correlation with the first principal component extracted from the APL inventory broken down by skills was the reading subscore. Similarly, the tests producing the strongest correlations with the first principal component on the

ITPA were the Auditory-Vocal Association Test (.96) and the Grammatic Closure Test (.92)—each of which involves filling in blanks in sentences.

Another source of reasoning that arrives at a similar result is to consider the content of the so-called "non-verbal" tasks. It turns out that the complex reasoning necessary to solve the Raven Progressive Matrices or Cattell's problems, or arithmetic problems, visual analogies, etc., almost certainly involves the deep structure of propositional forms that are primarily linguistic in nature. On the other hand, if we examine the content of the language tasks with which the "non-verbal" type of processing might be associated, we come against a blank wall. How, for example, could we reason that filling in blanks in a passage of prose about, say, a cat named Traveller, involves arithmetic computation? Or the solving of non-verbal analogies or abstract matrix puzzles? The so-called "non-verbal" tasks most certainly involve deep level language ability, but the normal processing of discourse does not seem to necessarily involve the solution of abstract visual puzzles or mathematical problems. Therefore, the content analysis of the tests supports the interpretation of the general factor as a language proficiency factor.

Yet another source of evidence on the question is theoretical reasoning regarding the nature of language proficiency. We should expect to be able to measure the global factor of language proficiency with a variety of language processing tasks. In fact, the correlations across tasks that are quite different at the surface should, under specifiable circumstances, roughly approximate the estimated reliabilities of the respective tasks. That is, if a written cloze test and an oral interview procedure are both tapping the same global factor of language proficiency, the respective correlations of those tests with the global factor in question should roughly equal the reliabilities of the several tests. This result is a logical necessity if the tests are really measuring the same underlying ability. Moreover, such a finding would support the view that the underlying ability is in fact rather unitary. To show that the ability in question is complex in the sense argued earlier by Thurstone and Thurstone (1941) and by many other theorists, would require the demonstration that stable specificities exist which can be associated with multiple unique factors.

The problem of demonstrating specificities (that is, reliable variance unique to particular tests) is compounded by the fact that in many of the original studies on the structure of intellect, the tests employed had remarkably low reliabilities to start with. While correlations across language processing tasks of many sorts can be expected to mount into the .7 to .9 range, correlations across many of the multifarious tasks used in previous investigations of "intelligence" in its many posited forms have usually been much lower. For instance, in the classic study by the Thurstone's, correlations across their 63 measures (3,969 in all) scarcely ever exceed .6 and are mostly below the level of .2. The significance of the relative magnitudes of the correlations in question can be appreciated by considering the fact that a .2 correlation indicates only a 4% variance overlap while a correlation of .8, say, indicates a variance overlap of 64%.

Although it can be argued under certain highly restrictive conditions that low correlations indicate that tests are validly tapping separate abilities, that is almost certainly not the case here. Here, the tests are quite probably unreliable to start with.

A test can only be said to be reliable if it correlates with itself. That is, if we repeat the measure, we should expect to get somewhat similar results. However, it is highly doubtful that tests such as writing as many words as you can that begin with the letter T and end with E, or trying to recall from rote which first names go with which last names in a previously presented list, or writing all the words you can think of that begin with D, or writing as many four letter words as you can that begin with B, or replacing missing letters in isolated words, are reliable. Although all of these tasks appeared in the set used by the Thurstones to investigate the "factorial structure of intelligence", they bear little or no resemblance to the sorts of things that people normally do with language. Further, they did not correlate strongly with anything—not even each other. Surprisingly, at least 42 of the 60 tests used by the Thurstones correlated more strongly with a single general factor than they did with the seven separate primary factors posited by them (see pp. 88 and 89 of their book). In fact, even the loadings of the tests supposed to be assessing the various primary factors (see p. 89) were generally quite weak (rarely above .6). Contrast these results with studies of language proficiency where the loadings on the general factor are often in .8 to .9 range.

In conclusion, let us briefly examine some of the most recent evidence in favor of a global factor of language proficiency from studies of second language learners. One of the most widely accepted claims of discrete point theory is that oral language proficiency can be divided up into various components. This claim has been tested carefully now several times over. In every single case the evidence shows that multiple oral scales of, for instance, pronunciation, vocabulary, grammar, fluency, and comprehension, are quite perfectly unitary within the limits of their respective reliabilities. The first careful study of this question came from a dissertation by Hinofotis (1976). She used the five scales from the Foreign Service Institute Oral Interview in conjunction with various other scores including scores on the Test of English as a Foreign Language. A principal components analysis of her data revealed a single language factor accounting fo no less than 65% of the total variance in the procedures studied (Oller and Hinofotis, 1979). Loadings of the various subscales of the FSI Interview on the global language factor were roughly equivalent to their reliabilities and there was no tendency whatsoever for tests aimed at the constructs of pronunciation, vocabulary, and grammar to correlate more strongly within categories than across categories. In other words, the various vocabulary scores obtained were no more strongly correlated with each other than they were with grammar scores, or measures aimed at other constructs.

More recently, Yorozuya and Oller (in press) have followed up on the Hinofotis research to test for the possible existence of a halo effect which might have been mistaken as a separate speaking factor in previous studies. Their results showed a substantial tendency for judges to assign similar ratings to diverse scales if they were scored on the same occasion. However, the data did not reveal any consistent specificity associated with any of the traditionally distinguished "components" of speaking skill. It would seem that ratings of "fluency" or of "vocabulary", etc., are in fact ratings of the same global factor of proficiency—perhaps it should be called "communicative effectiveness". When the halo effect was present the first principal component accounted for 89% of the total variance across four scales studied (grammar, vocabulary, fluency, and pronunciation), and when the halo effect was eliminated by having judges assign ratings to each scale on a different hearing of each interview, the first principal component accounted for 79% of the total variance. The halo effect, thus, seems to account for about 10% of the total variance in such scales. There was no stable specificity in any of the scales. In fact, the average specificity was less than 0 indicating that the scales were somewhat better measures of the global factor than of the separate constructs they were supposed to be measuring.

Additional data on this question are available from the correlations reported by Mullen (1979b) in a study of the reliability of oral ratings. In her research 98 foreign students at the university level were interviewed. Five rating scales were used (listening, pronunciation, fluency, grammar, and overall). She argued that each scale was contributing something unique to the evaluation of oral proficiency, but a principal components analysis based on the correlations she reported revealed the results shown in Table 2.

TABLE 2

A Principal Components Analysis of the Correlations Reported by Mullen (1979b) across Scales Aimed at Various Aspects of Speaking Proficiency of Second Language Learners

N = 98

Scales	Loadings on a General Factor, g	Proportion of Variance in Each Scale Explained by g
Listening	.930	.866
Pronunciation	.905	.820
Fluency	.926	.857
Grammar	.915	.836
Overall Rating	.980	.960

Whereas the "overall" scale was apparently the best all around measure of whatever was being measured by the various scales, all of them measure that same attribute to the extent of their reliability. In fact, the weakest correlation of any scale with a single proficiency factor was .905—accounting for 82% of the variance in the pronunciation scale. There is no reason to believe that any of the scales is contributing any unique reliable variance (i.e., specificity) to the evaluation.

Finally, a similar set of data on 117 foreign students tested in an essay format by Mullen (1979a) yields essentially the same conclusions. In this case examinees were asked to write an essay and their efforts were rated on five scales: structure, organization, quantity, vocabulary, and overall. Again Mullen concluded that each scale was contributing something unique to the evaluation, but a principal components analysis based on the correlations reported by her supports a somewhat different conclusion. See Table 3. As in the case of the oral scales, all of the scales for essay ratings loaded on a single component at remarkably high levels. The first component accounted for no less than 85.7% of the total variance and received loadings ranging from .87 to .98 from the various scales. To the extent that the scales are generating reliable variance, they seem to be measuring the same global factor of language proficiency.

TABLE 3

A Principal Components Analysis of the Correlations Reported by Mullen (1979a) across Scales Aimed at Various Aspects of Essay Performance of Second Language Learners
N = 117

Scales	Loadings on a General Factor, g	Proportion of Variance in Each Scale Explained by g
Structure	.906	.821
Organization	.936	.877
Quantity	.870	.757
Vocabulary	.931	.866
Overall Rating	.982	.965

Testing Panel

The Role of Standardized Tests at the School Level

Paul Rosier

Page High School, Page, AZ

Standardized Tests have come under increasing criticism in recent years. This criticism seems to have arisen as the awareness of the educational needs of minority children has developed. The validity of standardized tests has been and is being questioned, e.g., do these tests measure what they are supposed to measure? IQ tests have been the focus of greatest challenge. The construct intelligence quotient is difficult to define at best. Determining how to measure this construct has created voluminous controversy.

Standardized achievement tests are now coming under scrutiny. Again what is actually measured with standardized achievement tests? One aspect of these tests is the measurement of basic skills. Reading is a commonly measured skill. But critics have demonstrated that much more than reading skills are measured in such tests. Reading is not an isolated function. It is an aspect of language, and language is the means by which we define and describe our environment, and culture. Consequently, a reading test reflects not only reading skills, but language manipulation skills and environmental-cultural knowledge.

Standardized tests will reflect by their very nature, environment and culture. No test is language-free and no language is culture-free. Therefore, culture and language will be reflected in any test. In the case of standardized nationally-normed tests, the culture will be the one reflected by the majority—the middle class, and the language will be English.

Obviously, children who do not manipulate English well (i.e., those who have limited language structures and limited vocabularies and children who have environmental and cultural experiences different from the middle class) generally will not do well on standardized tests. This is the situation in which most of minority children in this country find themselves. They speak limited or non-standard English and their environment and culture are quite different from that of the middle class. These children will be at a disadvantage in taking such tests until they develop proficient English skills and acquire experiences similar to the middle class.

With these inherent weaknesses, what should be the role of standardized tests at the local school level where school populations include minority children? There is no absolute answer, but I will attempt to provide a plausible response by examining the following aspects of the question: (1) the use and abuse of standardized tests; (2) what standardized tests measure and (3) what comprises a comprehensive testing program.

1. The use and abuse of standardized tests

The most commonly used standardized tests administered in schools are achievement tests. Probably every student enrolled in either public or private school has taken at least one, and very likely many, batteries of standardized achievement tests. Testing can begin (and often does) in kindergarten with a set of standardized readiness tests.

Many states are now mandating standardized testing and/or are setting standards that require measurement by norm-referenced tests. The state of Arizona requires both the use of standardized tests and standards measured by norm-referenced tests. At the beginning of third grade all students in Arizona public schools are tested in reading. Presently the Stanford Achievement Tests are administered by local school districts state wide in October. Entering fifth grade students are required to take a standardized test in mathematics.

Arizona requires that all students graduating from eighth grade must read at a minimum level of 6.0 grade-level-equivalent years. Seniors, in order to graduate, must read at a minimal grade-level-equivalent of 9.0. Each local school district must file a plan with the state for determining the grade-level-equivalents of 6.0 and 9.0. The determining of whether or not students have achieved these levels requires the use of a locally normed or nationally normed measurement instrument.

Federal programs such as ESEA Title I often require evaluation procedures which incorporate some type of standardized testing. Usually this is a pre-post testing program in basic skills.

What effect does the use of standardized tests and state imposed standards have on minority children? I use Navajo children as the primary example since these are the children with whom I have worked for the past ten years. Navajo children on the average score as follows on standardized test achievement: At the third grade level these children score 1.0—1.5 grade-level-equivalent years below the national norm; at the sixth grade level Navajo children average 2.5—3.0 grade-level-equivalent years below the national norm. After twelve years of school Navajo children score on the average between 4.5—6.5 grade-level-equivalent years below the national norm.

Obviously, most Navajo children will not meet the state mandated standards for graduation from eighth grade and twelfth grade if the national norm is used as the standard. Many school districts serving Navajo children have determined their own norm and have developed an instrument locally to measure student achievement. This local norm often is well below the national norm.

The goal of the Arizona mandate was to increase the reading ability of students. Rather than attempting to improve the educational program to increase the reading ability of Navajo children, many districts have circumvented the mandate and have developed an evaluation system that allows the status quo to continue.

The use of standardized tests in Federal programs (Title I) reveals a most interesting situation. Programs first must demonstrate need. This is usually achieved by demonstrating with standardized tests that children score well below the national norm. Once funds are received then results must be demonstrated. If results are too great and children score at or near grade level, federal funds are terminated. The dilemma is obvious: improvement is needed, but not too much.

While conducting research, I encountered a situation that demonstrates abuse of standardized testing with Navajo children. The research involved two standardized tests used to measure reading achievement: (1) the Stanford Achievement Test (SAT): and (2) the Metropolitan Achievement Test (MAT). The total reading subtests of the two tests were quite different. The SAT subtest was composed of a subsection on phonics skills and a subsection on paragraph comprehension. The MAT total reading subtest was a combination of a word knowledge (vocabulary) section and a paragraph comprehension section. Children in four or eight schools used in the research study were tested on both instruments. The tests were administered to children in grades second through sixth. The average scores at each grade level in all four schools were lower on the MAT than the SAT. The word knowledge section of MAT was devastating to Navajo children. Although Navajo children in general scored well below the national norm on the SAT, the Bureau of Indian Affairs has chosen to use the MAT reservation wide for students' achievement evaluation. The reason for this choice seems apparent. This raises the question of the purpose of testing and the use of results.

The greatest abuse of standardized tests is the use of the results. This is particularly true for individual student test scores. Test results are often used to predict expectations both formally and informally. Children who score similarly on standardized tests are often grouped together for instruction and such groupings can be an effective method. However, minority children all too often are grouped together and are labeled as the "dull" groups. Teachers review the standardized test scores and often interpret test scores as aptitude. Children score low, therefore they must be slow or have little aptitude. Consequently, teachers expect little of these children and the self-fulfilling prophecy occurs: children do little.

2. What standardized tests measure

Standardized achievement tests are verbal instruments that require reading and symbol manipulation. No matter what skills or concepts are included in the test, language manipulation is involved in all areas except for arithmetic computation. Consequently, standardized achievement tests measure not only specific skills and concepts but how well a student manipulates standard English. The Gates McGinite Reading Test, Survey E, Form 2 provides several examples of how the ability to manipulate English syntax is essential to successful performance on the test. Most of the comprehensive questions provide answers which are contextually cued. Native speakers immediately eliminate those foils that are contextually inappropriate, but limited-English speakers may not. To illustrate, here is an example from the test: "Mother Goose rhymes seem like innocuous jingles, designed only to entertain small children. But many began as political satire using the nicknames of 27 or court 28 and making fun of their activities." The possible answers for No. 27 are (1) infant; (2) political; (3) stars; (4) rhymes; (5) animals. Contextually or structurally inappropriate foils are numbers one, four and five. The entire question implies a knowledge of Mother Goose rhymes which many minority children do not experience. Middle class cultural knowledge also is important if children are to have reasonable opportunities to demonstrate their skills on such tests.

Hence, the use of standardized achievement tests with minority children often reflects more how well they manipulate standard English and middle class cultural concepts, than how well these tests measure skills and specific concepts learned. To a large degree then achievement tests reflect how well minority children manipulate standard English and middle class cultural concepts in relation to middle-class students. These tests by no means reflect the actual ability or potential of these children.

The above information can be useful in setting goals and in designing programs. Minority children must compete in an English speaking, middle class world. In a non-moral way minority children start school behind middle-class children, i.e., they often speak limited or non-standard English and have had experiences quite different from those on which school curricula are based. In essence these children must learn more and learn faster to catch up (the Bereiter-Engelmann dilemma). Standardized achievement tests can be useful instruments in assessing the success of programs designed to catch these students up with native English speaking, middle-class students.

3. A compehensive testing program

Standardized achievement tests are not going to disappear; in fact the use of these tests may be on the increase. Standardized achievement tests can be useful when integrated into a comprehensive evaluation system.

Standardized tests alone seem to be of relatively little value. They often do not measure what has been taught in a given classroom or grade level. These tests do not serve well as diagnostic instruments because they are designed to measure the general rather than the specific. They do provide some information as to how students in a given area compare to other students throughout the country.

Used as one component of a comprehensive evaluation system standardized tests can provide useful information. An effective evaluation system is an out-growth of a well planned and designed curricular program. Objectives and goals are developed for specific skills and concepts to be taught in the curriculum. Evaluation results are used to assess whether goals and objectives are met and to determine program direction.

A combination of a locally developed criterion—referenced testing (CRT) program and norm-referenced standardized tests can provide a comprehensive and effective evaluation system. The CRT program can assess the extent to which curricular goals and objectives are met. This testing program evaluates how effective classroom instruction is. What is supposedly taught in the classroom is measured.

The overall goal of any educational program is to develop student skills and abilities so that students can become competent citizens. Minority children have the right to such an education. The goals and objectives of the school curriculum must be designed to assist these childen in developing competencies equivalent to those of middle class students. The standardized achievement tests component of the evaluation system can be used to assess the progress minority students, as a group in a given school or grade, are making toward that goal. The data gathered from such tests may be used to review and alter curricular objectives.

The Office of Education has suggested that there be a moritorium on the use of standardized tests. I believe this to be an undesirable move. On the Navajo reservation some schools have decided not to use standardized tests. The Bureau of Indian Affairs has substituted local norms for national norms for the SAT and MAT. The results of these actions have not improved the competency levels of Navajo children. All that has been accomplished is that mediocrity has become the accepted standard.

With all their weaknesses standardized tests can be valuable. These tests provide data which describe how well students, through the medium of English, achieve in comparison to students throughout the country. How the information from these tests is used determines their value. If the information is used in isolation and for individual students, it is of seemingly little use. If it is used in a comprehensive evaluation system, the information can provide program direction.

Standardized tests are not going to evaporate. The challenge educators face is making positive use of these tests. Certainly standardized tests can be better designed. School personnel should carefully examine a wide variety of such tests before the selection of a specific test is made. The test should be one that provides information needed for the norm-referenced component of the school's comprehensive evaluation system. We, the school personnel, to a large degree decide whether the data gathered from standardized tests are useful and valuable or whether they are abusive and detrimental. It is in the role as decision makers that we must use standardized tests for no more or less than they can assist us in educational decisions.

Testing Panel

Three Kinds of Bias in Standardized Testing

Roger W. Shuy

Georgetown University,
and The Center For Applied Linguistics

Any examination of standardized testing would do well to attend to T. S. Eliot's three probing questions: "Where is the learning we have lost in information? Where is the understanding we have lost in knowledge? Where is the life we have lost in living?" The notions that learning can be measured by means of retained information and that understanding is reflected by knowledge have been challenged by wise people for centuries. At best, testing which is standardized to large numbers of takers is guilty of assessing information and knowledge rather than the learning and understanding which Eliot so highly valued. Information and knowledge are the *what's* of education. Learning and understanding are the *why's*.

But even assuming that it is valuable for education to try to measure knowledge or information, several major problems obtrude. What is worth measuring? What is an appropriate sample of the larger body of knowledge which represents that knowledge? What gets in the way of an accurate measure of such information? The latter issue—that which interferes with such measurement—often goes under the label, bias.

Bias in standardized testing exists at many levels. We usually think of bias in terms of cultural bias, but it is difficult to think of the category of test bias without considering the bias of the general public and the bias of the disciplines as well. This paper will discuss these three areas of bias as they relate to standardized testing today.

Education, as every teacher knows, suffers from the attitudes of the general public which thinks it knows all there is to know about education and that special thought, study or experience with education is not needed to qualify one to make grandiose statements about it. Since educators appear to be under siege, careless interpretation of what tests actually do can be damaging to our profession. There exists today an industrial/business mind-set which causes the general public to believe that we actually know how to measure such things as intelligence, language proficiency, reading, etc. This mind-set, which appears to be modelled on simpler testing of how well automobiles can run or how widespread a television program is watched, leads the general public to believe in the notion of the standardizable testing of human learning and to believe that we already know all we need to know in order to do it. What all educators know is that human children are not like automobiles or television programs. Nor is learning that finite or controlable.

This bias of the general public was made clear to me in court litigation concerning the Aspira Consent Decree in New York. It seemed perfectly obvious to the litigants, to the defendants and to the judge that what was needed in New York was a test which would clearly determine in which language, Spanish or English, the New York City Puerto Rican children could most effectively participate in the classroom (reading, writing, listening and speaking). This having been agreed on, all that remained to be done was either to select such a test from among those available or to make up a new one to suit their needs. The simplicity of this case was apparent to all. The matter was referred to the appropriate division of the New York City schools, the research department, which did not find the issue quite so clear. In New York, as in other places, pressure from court orders, legislation, community groups and the needs of school children have demanded immediate response, response before an adequate body of information has been developed. Furthermore, what is to be tested is a complex issue. On one level, proficiency in English must be assessed as must proficiency in the other language. The other language differs from community to community. On another level of complexity is the issue of what to test in all the languages. Which aspects of the language system must be included in a test of proficiency? Which can be inferred from the results of other aspects that are tested? What aspects of the language systems can be safely ignored? Currently, tests stress phonological, morphological and lexical items. Some involve syntax. None capture adequately the pragmatic and semantic aspects of language that, on the face of it, are important elements of proficiency in a language. Nor has there been much recognition of the fact that we really do not know much about how to test these more crucial areas of measurement in order to satisfy the rather nice wording of the Aspira Consent Decree ("to effectively participate in the classroom"). The business industry model does not permit it.

In testing, as in most other aspects of education, the business/industry model is clearly at work. There is an assumption that we already know how to do the task and that all we need to do is measure it. A further extension of the business/industry influence can be seen in the immediacy of the request. Ironically, we often accuse the younger generation of wanting their products now—in fact we call our children the *now* or *me* generation. It appears that they come by this naturally. The general public tends to be unwilling to defer gratification (a term we once saved for the underprivileged). It is, of course, an error to think that we already have a test to satisfy the needs specified by the Aspira Consent Decree and it is equally erroneous to think that we can build one without considerable research, thought and time.

There was a time when one could count on philanthropic foundations to see through the superficiality of the business/industry model of education. As their resources diminish, however, even these philanthropic groups are being forced to move toward the need for an immediate, visible, product pay-off. The point I am trying to make here is that however troublesome cultural bias may be, it is probably no more damaging than the ignorance of the issues caused by public bias. Education is not highly regarded in this country, as evidenced by the fact that most education legislation is hinged to social, political or economic concerns. Seldom, if ever, is learning considered for its own sake. If education is so invisible to the general public, how much more invisible is language? Here we have a field in which newscasters can write best-selling books in a discipline whose existence they do not even recognize.[1] Nor does the general public hesitate to offer sharp opinions about bilingualism, language standards or clear writing, despite the fact that there is no general scientific agreement about what these concepts mean. Public bias has been little recognized in education as a whole, and in language education in particular. In my opinion it is a dangerous enemy of the abilities which must be measured in standardized testing of any kind. Any public which thinks we know how to carry out such testing while we really don't, will continue to accept specious test results.

A second broad form of bias which affects standardized testing comes from the disciplines which are involved. Somewhere, somehow, the uneasiness which most teachers have with the process of testing was exacerbated by the quantification leading to great disciplinary confusions. The construction of any testing instrument must consist of three components: what is measured; how it is measured; and what the results mean. Just as the business/industry model has influenced public bias, so has it surfaced in disciplinary bias. The delivery system is in danger of becoming the message. In an honest effort to improve our methodology of how to measure, we run the risk of losing sight of what is being measured. Put another way, our effort to find reliability can interfere with our notion of content validity.

A classic example of this problem can be seen in the Scholastic Aptitude Test (SAT). As you know, reliability in testing is the degree of consistency between two measures of the same thing. This is probably what the College Entrance Examination Board does best. The SAT compares a test-taker's capacities with those of his peers and with students who took the test before. The test is also correlated with the students' grades in high school and, it is claimed, over the years, it has actually improved at doing the only thing it professes to do, to predict success in the first year of college. Likewise, different forms of the test are measured against each other for reliability.

Reliability is, without doubt, a good thing but by itself it does not really tell us very much because any test can be technically reliable and still fail to measure what it intends to measure. The attractiveness of reliability is attached to the fact that it can be quantified, thus giving the air of scientific respectability. Psychometrics has perfected the area of reliability, and for this we can be thankful, but the more central issue involves the question of the validity of what is measured and the effectiveness of the test to measure what it says it measures. Since the SAT claims not to be an achievement test but, rather, a predictive measure-for-success-in-college instrument, content validity becomes a cloudy notion. The SAT is called an intelligence test by some since the difference between measurements of intelligence and measurements of aptitude are slight at best.[2]

---

[1] In a recent conversation with news commentator, Edwin Newman, for example, I explained that I work in the area of linguistics and literacy. "There is no relationship between linguistics and literacy," Newman responded.

Thus the SAT measures verbal and mathematical factors rather than all of the things one might want to call intelligence (provided we could ever agree on what these things are). During the history of the development of intelligence tests, the term intelligence was thought to be unchanging and innate. To avoid the implications of innateness, test developers have used the term aptitude. Since these tests are used primarily to predict school success, the term scholastic was added. One could make the case that the SAT is really an intelligence test and that, from what little is known about intelligence or how to measure it, the SAT is invalid. If one takes the position that the SAT is an aptitude test, however, the test is subject to analysis of why such group aptitude tests succeed at predicting school achievement.

In theory it is possible to distinguish between aptitude and achievement. One is an innate property and the other is accomplishment. Once content validity has been determined, it is possible to measure achievement. How to get at aptitude, however, is not so clear. The SAT selected the mathematical and verbal areas as a locus for measurement. Exactly how this decision separates achievement possibilities from aptitude remains very, very unclear. All children have knowledge, memories and feelings about the subjects being measured. Furthermore, distinguishing among SAT verbal aptitude items and those items in verbal abilities on achievement tests is virtually impossible. The distinction between aptitude and achievement, though possible in theory, is impossible in operational practice. What is more, ability appears to increase as achievement is acquired. Why is it, then, that group aptitude tests can predict achievement? It is highly probable that they are really achievement tests.

It has been necessary to follow this reasoning in order to address the issue of content validity. And it is in the area of content validity that the test making industry and the other academic disciplines fall away from each other. Once verbal and mathematical achievement are recognized as integral to any assessment of the SAT, we must ask the question—how do we know that the SAT measures the right things? Willard Wirtz headed the blue ribbon panel which spent two years studying the decline in SAT scores. The panel did *not* address this question although, to its credit, it did recognize it as a legitimate problem: "We have accepted, for purposes of this inquiry and report, the traditional value base of the SAT's validity. . . ." (p. 11).

---

[2] The two basic (and relatively minor) differences between intelligence and aptitude tests are:

a. That aptitude tests measure specific factors rather than general ones;
b. That intelligence tests measure general factors rather than specific ones.

Perhaps one of the most significant statements of the panel report grew out of a discussion of why scores on the Achievement Tests (which may also be elected by SAT takers) did not parallel SAT score declines. Scores on tests in English composition, French, Spanish, biology, chemistry and physics actually increased while other subjects reveal only small declines. The panel lamented that it had not been able to analyze this phenomenon fully and concluded:

> "It is conceivably important that the College Board make much larger use of outside committees in connection with the Achievement Tests than with the SAT: the counsel sought is in the one case from experts in the particular disciplines, in the other from psychometricians and psychologists." (P. 23)

This criticism is at the heart of an extremely important test bias—the bias between disciplines. This criticism appears to be one of the central issues in accounting for the SAT score decline. The SAT is actually a disguised achievement test which, in order to be valid, must better address what really matters in the two content fields which it measures. Like most objective tests, it is a discrete point measure. That is, certain items are selected from an inventory of all possible knowledge in order to represent that larger knowledge in as few questions as possible. Only those who know the content of the fields being measured have the key to content validity. Test makers create items to test only after content validity has been established.

This lengthy analysis of the SAT was meant to serve as a touchstone for the issue of disciplinary bias. In general, many psychometricians feel the need to develop tests which discriminate between students. A certain number should get high marks and a certain number should get low marks. There is often no particular interest in why these people are so distributed. Therefore, when the New York State reading test is given to tenth graders, it is estimated that some 7,000 children will fail in New York City alone. What other disciplines are asking is, why this should be so. What is there in these children which will cause them to fail? And if we can show, on a standardized test, that they have failed, why don't we take these areas of failure and highlight them in their school curriculum? Those who make up these tests are not often interested in this issue. Those of us in the disciplines of language and linguistics, along with many others, care less on the whole about reliability and pupil distribution than we do about validity or about the question of why the pupils distribute as they do and learning rather than placement.

Quantitative analysis in linguistics works best at the individual level.[3] When we go to compare one individual or group with another individual or group we fall into treacherous waters for we face the same problem that every standardized test faces: the question of thresholds.

---

[3] See, for example, Roger W. Shuy, "Quantitative Analysis: A Case for and Some Warnings Against," in *Anthropology and Education Quarterly*, VII, 2 (May, 1977) 78-82.

Whenever the threshold of acceptability is decided, it is an arbitrary decision. In the case of the Aspira Consent Decree, the ultimate question was "What is the cut-off score at which children are sent either to bilingual or to mainstream classes?" For those who have never witnessed the process of establishing such a threshold, let me explain how New York City did it. First, they developed a test mostly of English morphology, phonology and vocabulary for Spanish speakers. Then they gave that test to 15,000 New York Anglos (a term which, curiously enough, included Blacks). Then they took the score which separated the top 90% from the bottom 10% of the Anglos who took that test as the cut-off point to be used to separate Puerto Ricans into bilingual and mainstream classes. The federal district judge who presided over this case then asked the New York City schools where that 90%/10% cut off point came from. With that question, Judge Frankel asked one of the most penetrating and critical questions in all of standardized testing. The answer, of course, was that it was arbitrary. The exact words were: "It was an administrative decision".

Bias of disciplines, then, is an important consideration in the examination of standardized testing. I cannot speak for all disciplines in this, of course, but I can safely say that the bias of linguistics is to the validity of the content being measured. This is most important of all to us. Reliability, a crucial issue to other disciplines, is a fine thing but only after validity is clearly noted. It is unfortunate indeed that the content fields have not yet progressed to the point at which we can state exactly what it is we want to measure in terms so that the science of psychometrics can be best utilized. Lacking this, test makers tend to proceed with what they have. And in a sense they cannot be blamed for the weaknesses of our field. But it is our obligation to call attention to these weaknesses and to work toward correcting them.

The Wirtz Committee dismissed the issue of potential cultural bias in the SAT by observing that a definitive analysis of cultural bias is virtually impossible. It appears, however, that a definite analysis of cultural bias is no more difficult than definitive analysis of many other educational concepts. Test questions of any type involve matching the assumptions, values and presuppositions of the test maker against those of the test taker. An understanding of the basic principles of language and culture make the identification of potential cultural bias available to those who care to find it.

The major issue of cultural bias is similar to the issue of so-called language dominance tests. Such tests have all the problems of proficiency tests plus a problem which derives from a very uninformed notion of dominance as a static, context free characteristic. Much of the exciting new developments in sociolinguistics, psychology, philosophy and anthropology are concerned with the effect of context on learning. Context can be defined as what the learners bring with them to the learning process as well as the setting in which the learning takes place. Contrastive assumptions, values, presuppositions and setting in testing all contribute to the question of why students are spread across the distribution curve. These factors all can be examined, thanks to recent developments in speech act theory and communicative competence. The ways in which different cultures share these assumptions, values and presuppositions in different settings provide the key to cultural perspective. When this perspective does not match that of the test question, culture bias can be identified. Thus, what has been assumed to be virtually impossible to identify in the past, now becomes a researchable question.

This discussion of the three kinds of bias, public, disciplinary and cultural, is by no means definitive. Nor has the field of testing really addressed them satisfactorily. The exposition here is intended more as an agenda than as an answer. T. S. Eliot's questions may well serve as a general guide to our thinking about testing for we are certain to be better off if we can learn to assess understanding and learning rather than information and knowledge. But even in our efforts at the latter we have overlooked the powerful biasses of the expectations of the general public, of the disciplinary perspectives involved in producing and interpreting the tests and the mismatch of the assumptions, values and presuppositions of the test makers and test takers. All are biasses of one sort or another. All can cause us to lose sight of life in the process of living.

Methods Panel

An Exploration of New Trends in Second Language Teaching: the Silent Way, Suggestopaedia, Counseling Learning-Community Language Learning

Wilga M. Rivers

Harvard University

At the beginning of *Teaching Foreign-Language Skills* I took you on a tour of four classrooms, all engaged in different activities, but all devoted to the *teaching* of a language (Rivers 1968: 1-7). Now I think it's time for us to take a tour of our Foreign Language Institute in 1979 and visit three classrooms, all devoted to the *learning* of a language.

In Classroom A, there is a long silence; the teacher looks encouraging; a student constructs a sentence; the teacher nods, changes the positions of some rods, and gestures to another student who, after a moment of contemplation, puts together an utterance and is encouraged by the teacher's reassuring smile.

In Classroom B, a small group is seated in a circle with a tape recorder in the center—the recorder is there not to give but to receive language material. The teacher is not obviously the focus of this scene, but with some aides is outside the circle, waiting for students to take the initiative. One student is speaking, occasionally seeking a word from an aide outside the circle. Another student cuts in, rather aggressively, refusing to accept what the first student has just said. The other students murmur further objections, seeking help for stronger expressions of their displeasure from the teacher or from an aide. The latter supplies exactly what is requested in a low tone without intervening in the exchange. All of these utterances are being recorded for later analysis by the group in discussion with the teacher.

In Classroom C, baroque music is playing. Soft light falls on the pastel walls and the students look pleasantly relaxed as they recline, breathing rhythmically, and listen to a dialogue being recited, alternately in a normal tone of voice, then in a persuasive whisper, and then triumphantly. The dialogue is quite long, so we leave before the end of the rendition.

What can these three ways of approaching the language-learning task have in common, and how is it that they have all come to the fore in this particular decade? We may note that they are not newly forged. Gattegno's Silent Way in Classroom A was clearly described in the first edition of *Teaching Foreign Languages in Schools: The Silent Way* which appeared in 1963. At that time, Gattegno stated that it was based on 35 years of language-teaching experimentation in 14 countries on 5 continents. Curran's Community Language Learning in Classroom B was first described in the *Bulletin* of the Menninger Clinic in 1961 (and I discussed it at length in *The Psychologist and the Foreign-Language Teacher* in 1964). Lozanov's Suggestopaedia was the subject of experimentation in Bulgaria in the 60's although detailed accounts of it did not appear in the United States until 1970 (Ostrunder and Schroeder, 1970).

The reasons why these approaches have suddenly become important to us can be found in particular emphases which have evolved in foreign language teaching in the 70's. In the grammar-translation period (prior to World War II in most languages and still continuing in English in many parts of the world today) the emphasis was on "*Why* teach a modern language". The justification, for reasons of respectability, was based on intellectual rigor and cultural enrichment, with the result that the teaching of modern languages became as much like the teaching of classical languages as possible—with emphasis on the understanding of a formal grammatical system and ability to read literature and philosophy.

Modern language teachers then passed through the *What* and *How* of the audiolingual era, asking such questions as What is language? and How do people learn languages? The structuralists believed they knew what language was. It was "a system of arbitrary vocal symbols by means of which the members of a society interact in terms of their total culture," as Trager expressed it; (Trager 1949:4) and the behaviorists knew how it was learned: by acquiring language habits through reinforcement (or confirmation of their efficacy). This emphasis on what and how was continued during the transformational-generative period, when the so-called cognitive-code approach was much discussed (though rarely described). Language we now knew was an innate structure and our language-learners through their innate language-acquisition device were hypothesizing about the form of the grammar to which they were attending and matching it against their innate knowledge of potential grammars of a human language. Despite the theory, in the cognitive-code approach as proposed in the literature, students had little opportunity to hypothesize about the form of the grammar. Since language use was considered to be "rule-governed behavior" (a term largely misunderstood) it behooved students to understand and apply the rules (Rivers 1976:11). We were back to a deductive presentation of the grammar, with conscious practice in the application of rules before authentic passages of discourse were encountered (Chastain 1976:156-157).

But with the 70's came renewed emphasis on the individual. *Who* are our language learners? How do *individuals* learn? What are their *personal* learning strategies? Second-language teachers became wary of presuming they could *teach* a language and began to seek ways in which students could be given opportunities, situations, and time to *learn* as their individual proclivities permitted.

The three approaches we are to discuss this afternoon all fit into this new mood of emphasis on the individual and personal learning strategies. Each one tries to give the student room and time to learn with as little intrusion of the teacher into the learning process as possible. Gattegno (1972:xii), for instance, speaks of "techniques which made it possible for the teacher to say less and less as the lessons advanced, while the pupils were saying more and more and using their own inner criteria." He speaks of throwing "the learner upon himself" (Gattegno 1972:ix). Although all three approaches structure the learning situation, and structure it in very specific ways based on theoretical convictions, they each envision the teacher's role as indirect. This applies even to Suggestopaedia where the teacher's role is highly structured, but the emphasis is still on allowing time and space for the students to use their language-learning capacities to absorb and assimilate the material.

What else do these three very different approaches share?

1. They all endeavor to involve the whole person of the student. Curran calls one of his books, *Counseling-Learning: A Whole Person-Model for Education* (1972). Gattegno considers breathing and kinesics of tremendous importance in speaking a new language, as well as the conscious application of the intellect. Lozanov tries to draw on the powers of the unconscious mind to assimilate positive suggestive factors in an act of communication or a pedagogical process. "The purpose of suggestopaedia," we are told, "is to put those 'mental reserves' to work by organizing a coherent system whereby the suggestive, emotional, more or less unconscious signals we receive, are coordinated" (Racle 1979:40). Curran draws these notions together when he speaks of learning as a "unified personality encounter" (1976:41).

2. All view the learning of a second language as quite different from the learning of the first. Gattegno calls it "radically different" (1972:xii). Clearly the Lozanov procedures make no attempt to reproduce child native-language learning situations, and Curran encourages students to draw freely on the native language in early attempts to communicate.

3. All three are inductive in the initial encounter of the student with the language. Curran's students learn through use, Gattegno's through inductive analysis of the teacher's (limited) output, and Lozanov's students assimilate structure in the dialogue, which they memorize but without conscious effort. All three provide explanation at a later point in the learning as students require it. Initially, however, all three present material and let the students do what they can with it. To quote Gattegno: "the synthetic, the intuitive, precedes the analytic" (1972:8).

4. All three approaches are non-corrective and give the student time, thus reducing the anxiety and tension of language learning. Corrections are supplied in a supportive way, as information rather than reproof, and as the student shows a need or desire for them. Gattegno makes a strong statement in this area "I do not correct learners," he says, "I only throw them back onto themselves to elaborate further their criteria and to use them more strongly" (1972:

31). Curran found that students wanted time. "The counselor had to be very sensitive . . . ," he noted, "to discern when he was really needed or when, on the contrary, persons wanted time to find, on their own, the required word or phrase" (1976:33). Lozanov's "concert" allows the students this time (Racle 1979:44).

5. Each of these approaches encourages active use of the new language in communicative situations from the very beginning. Curran's clients try to express their own feelings and opinions with the help of the counselor, or an aide, from the first moment. Gattegno's communication situations are structured by the teacher with rods, but the students are expected to convey information about these situations which they have themselves put together in the new language. Lozanov's students act out the material of the dialogue as soon as it has been assimilated and, by the structuring of the dialogue material, are "thrust directly into a communication situation (Racle 1979:46).

6. Finally, we may note that each of these approaches tries to create a community feeling of "all pulling together" (Curran 1976:1,7), which decreases inhibitions against expressing oneself in a language of which one still knows very little.

Each of these six factors can be reproduced in other approaches in a normal classroom: involving the whole person of the student (affective, cognitive, and physical); an inductive approach to language material; a non-corrective, encouraging attitude on the part of the teacher; active use of the language in communication as soon as possible; and the building of an esprit de corps where students begin to care for each other's progress and form a language-learning community. These are aspects of the language learning/teaching enterprise which have been urged on teachers for some years. Perhaps it is time now to do, not just to discuss.

Methods Panel

Explorations of New Trends in Language Learning: The Silent Way

Mary Hines
La Guardia Community College

When I realized I had fifteen minutes to explore the Silent Way, my first thought was to be silent. Then, time reminded me of watches—the Swiss watch with seventeen jewels and moving parts and the digital watch. Digital watches from Texas have little to do with watches from Switzerland. The two cannot be compared with the same criteria. One cannot criticize a digital watch for not having hands or a spring. One can criticize it for not keeping time right.

And so it is with the Silent Way. The method is a revolution and cannot be compared easily with familiar methodologies. It can be contrasted with familiar methodologies and it can be analyzed, something Earl Stevick does effectively in his book *Memory, Meaning and Method* (Stevick 1976). For details of the method I refer you to that book and Caleb Gattegno's own works *Teaching Foreign Languages: the Silent Way* and *The Common Sense of Foreign Language Teaching* (Gattegno 1963; 1976) and Fanselow's review of Gattegno's work in the TESOL *Quarterly*; Fanselow 1977).

The Silent Way is a revolution—a dramatic change. At the same time, it is a radical return to ancient, honorable theories of man, the nature of man, the nature of learning, the purpose of learning.

There is a mystique about the Silent Way, unfortunately, with its key words: rods, fidels, ogdens, inner criteria, awareness. Note the range of those words—the concrete rod to the abstract awareness. It is the last that I want to explore today.

I think there is this mystique about the Silent Way because Dr. Gattegno doesn't identify or relate his work explicitly with the tradition a reader recognizes inherent in it. The Silent Way is new. The Silent Way is old. The method is new. The philosophy behind it is old. Very old.

The sub-title of Gattegno's first book on the Silent Way is the "subordination of teaching to learning," a phrase that leads us back deep into our philosophical heritage.

Socrates considered himself a spiritual mid-wife—the teacher who could not teach, but who could help another give birth to his soul. Plato immortalized this Socrates, the inspiring questioner, one who provoked others to know that they did not know and thus to join the thoughtful search for self. (*The Republic*: trans. F. M. Cornford 1945) Centuries later, the French philosopher, Montaigne, reflected on the ancients and made a distinction between learning and study and identified study with education, a leading out of the self (Montaigne, *Selected Essays*, trans. C. Cotton and W. Hazlitt, 1949). Montaigne wrote that teaching and learning might impart knowledge, whereas study led to understanding; things known were made one's own and became a part of one's judgment. Education was a continuous heightening of consciousness, an increasing sharpening of judgment. Such is a description of awareness.

Examining the words of Plato, Robert McClintock (1971) notes that teachers "could not fruitfully instruct those who would not teach themselves, who would only respond passively to the most convenient appearance;" the most teachers could do was to convert inert souls to active study. "This theory of learning has sunk deep into our philosophical heritage, but it has not fared well in practice." It has not fared well. But, Dr. Gattegno reminds us that students must assume the responsibility for their own learning.

Dr. Gattegno is a classicist. He's also a Romanticist. Listen to his words . . . "It is important to think of language learning as a recovery of the innocence of the self." The Silent Way is Socratic. It is reminiscent of Rousseau.

Socrates, in his questioning, was nurturing not only the process of thinking; he was nurturing a faculty that grows with use, a faculty whose natural aim was the discovery of truth about the self. The distinction, again, of Montaigne. Learning provided tools. Study, education, was the search for truth. All of this is edifying. And we need to be edified.

Dr. Gattegno reflects on truth. He also urges teachers to allow students to establish an inner criterion for what is right and wrong, correct or incorrect about the language, and, one suspects, inner criteria are closer to Plato's cave than to interlanguage.

The philosophy I see inherent in the Silent Way places the school solidly in a sound, profound tradition of humanism. There are controversies, however. The teacher so respects the student, for instance, he expresses neither approval or disapproval. The responsibility for learning is the student's. So is the credit. But, while the subordination of teaching to learning is probably the greatest contribution the Silent Way has given us, this refusal to give either approval or disapproval is probably the most controversial aspect of it.

The other controversy is about the use of silence itself. Now, we all would like our students to learn English as efficiently as possible. The Silent Way is one way to learn. There are others. For the degree to which one follows the Silent Way in practice stems from one's reaction to silence. Encounters with silent bring peace, anxiety, tension, communion, reflection. Teachers who have experienced the Silent Way have shared their reaction to it. For one, it was a painful experience at first. Her insecurity, her defensiveness, came through. Another found herself having to deal with her own frustration, her nervousness, as a result of being silent. Both finally felt unburdened and found their silence allowed them freedom to become aware of students, the learners.

Other teachers, however, emerge from a Silent Way weekend saying they have never witnessed such a mechanical, impersonal, punishing class. For, the students, too, are encountering silence. The tension built up when the teacher is silent is frustrating for some students, and, just as some teachers say this method is impersonal, punishing, so do some students react. When an answer is not forthcoming from the teacher, they become angry. In some cases, the experience is so punishing, students leave. Many others, however, pass through the experience and thrive on it. They are happy to have a teacher step out of the way and let them take charge. Silence evokes tension. It evokes reflection.

Whether silence leads to tension, which explodes in frustration, or to reflection and creative exploring depends, possibly, on the teacher, but, more likely, on the student, his personality and his learning style. Yet it is here that the Silent Way is silent. I see no word on the idiosyncratic ways of learning. We are given the observation that language is not deduced but felt. Now, I may be trying to compare Swiss watches with digital watches to look for allowances for individual learning styles in a school that is based on the one nature of man, but with the controversy about the giving or withholding of approval or disapproval, it is a question.

In sum, the Silent Way is a multi-colored tapestry—as colorful as the rods—of philosophy and psychology and pedagogy. Its theory brings us back to the finest traditions of education. Its practice is unconventional. New. Old. Classical. Romantic. Silent. Paradoxes . . .

The paradoxes I see in the Silent Way remind me of an old Russian story. And I don't think I do violence to the tale when I associate language with life. Not when I think of Socrates. . . . It is the story of a rabbi who lived in a small Russian village. He spent his life searching for the meaning of life. He despaired of an answer. Then one night he had a dream. And in the dream he was told to go to Kiev. And in Kiev he would find a soldier, a sentinel, on the bridge over the river. The sentinel would have the answer for him. And so the rabbi sold all his belongings and saved for the trip to Kiev. And he travelled to Kiev. When he reached Kiev, he ran to the bridge and there he found a sentinel. "Thank god, you're here," he said. "I had a dream the other night. And the dream told me you would have an answer for me." The sentinel replied, "Oh, you should never pay attention to dreams. I had one myself the other night and in it a rabbi came up to me asking for the answer to life. I told him the answer was back in his home, behind his fireplace." And so the rabbi rushed back to his small village and in his cottage, behind the fireplace, he found the answer he was looking for.

For me, the story means that the spiritual treasure is there, with you, in the heart. But you have to go somewhere else, to another teacher, outside your tradition, to find the treasure. To find yourself, you must go to a stranger.

The stranger, for some in our field, may be Dr. Gattegno. For others it may be someone else but the important thing is that there be Socratic midwives to provoke us to know that we do not know and that we must continue on the thoughtful search. We are journeying. Let's continue to explore.

Methods Panel

A Review of Suggestopedy

Thomas Scovel
University of Pittsburgh

One of the many reasons I find attending the annual convention of our organization such a delightful experience is the opportunity to be exposed to the new: to meet new people and new faces from our evergrowing circle of professional colleagues, and, most important of all, to encounter new ideas and new theories, because they, and they alone, allow us to turn from the little we now know and look forward with an open but critical mind to the vast unknown that confronts each one of us. It gives me great pleasure, therefore, to see that the convention chairman, Carlos Yorio, has, in typical wisdom, arranged this panel as a plenary session with the specific goal of evaluating three of the most popular new methods of language teaching for you. It is a personal delight and honor for me to be able to come literally halfway around the earth, from the world's oldest continuing civilization—China, to this marvelous young infant of a city—Boston, to participate in this panel, and, if the supporters of suggestopedy are unfortunate enough to interpret what I am about to say concerning the Lozanov method as a personal affront, they at least will probably be just as happy to know that at the convention's close, I am flying right back to the other side of the globe!

William James once wrote that we should be tender-minded in the face of new ideas, but tough-minded in ever accepting them. I have already alluded to the importance of the former: let me sketch briefly why I believe we as language teachers and language researchers should be tough-minded in our acceptance, especially when we are given the weighty responsibility of speaking to hundreds of our colleagues at a plenary session. It is my belief that any assessment of the potential contribution to our field of a new theory, method, or technique should begin with the direct claims (or absence of claims) that the author of the innovation might make about its potential relevance to foreign language pedagogy. Because Georgi Lozanov, the founder of suggestopedy, and his followers do indeed make powerful claims about the enormous benefits that suggestopedy can effect on second language learning progress, it behooves us as fair-minded language teachers to take Lozanov seriously and give suggestopedy a thorough and attentive review. Let me summarize for you, therefore, a few of the claims that have been made, and then proceed to assess them from both a theoretical and practical viewpoint. The information upon which this brief critique is based derives from an intensive reading of the first English translation of Lozanov's *Suggestology and Outlines of Suggestopedy*, a review of several articles in popular magazines and professional journals, and attendance at a half-day session on suggestopedy given the fall of '78 at the annual convention of the TESL association of Ontario. Throughout this review, I have attempted to apply my tough-minded experiences as an applied psycholinguist at the University of Pittsburgh and those at my present position teaching EFL in northern China.

The claims begin with Lozanov himself. Using suggestopedy as an aid to the memorization of vocabulary items in foreign language classes, a teacher is guaranteed the following results in a 24-day course with four lessons a day—given at any time day or night and with no homework assignments: 1) 90% retention of a 2,000 word vocabulary, 2) the ability to "speak within the framework of the whole essential grammar," and 3), perhaps most astounding of all, to be able to read any text (Lozanov 1978:321-322). Finally, in perhaps his strongest claim, Lozanov proposes that his theory has a place, not only in psychotherapy and pedagogy, the two fields for which it was originally conceived, but in literature, music, sports, advertising, and commerce. "In fact," he concludes with gusto, "there is no sector of public life where suggestopedy could not be useful" (Lozanov 1978:3)!

His followers in North America have not lost any of Lozanov's zeal in promoting what they refer to as suggestopedia, a synonym for suggestopedy. Bordon and Schuster (1976), writing in the *Journal of Suggestive-Accelerative Learning and Teaching*, claim that learning is 2.5 times better with suggestopedy. In fairness, this is an enormously modest claim compared to a report issued by the Suggestopedy Research Group, established by Lozanov in his native Bulgaria, which enthusiastically stated, "As seen from the results obtained in the experimental groups, memorization in learning by the suggestopedic method is accelerated 25 times over that in learning by conventional methods" (Lozanov 1978:27). When such results are publicized in the popular media, the impression of a major revolution in language teaching is immediately intimated, especially when *Parade* magazine, with an estimated readership in North America of about 40 million, ran a laundatory story in the spring of '78 with quotes, such as the following from Dr. Gabriel Racle, who had an opportunity to publicize the Lozanov method at a plenary panel at last year's TESOL convention.

> The approach benefits the whole person. Not only do psychosomatic problems often disappear (headaches, stomach cramps, anxiety), but we see his personality develop; he becomes more self-assured, more creative, and expresses himself more easily. (Miele 1978:14)

What I found equally disturbing, however, about the current state of suggestopedy in North America is not solely that unverified claims such as these are continually bandied about as fact, but that individuals and organizations who may or may not have a thorough, professional training in foreign language teaching are making good money largely on the basis of these extravagant claims. As just one example, the Lozanov Learning Institute, established in Washington, D.C. and reputedly officially licensed to promote suggestopedy in the United States, was deluged with over 11,000 inquiries about the method after the *Parade* article last year. They were willing, in a form letter sent to me at Pittsburgh, to provide a 2½-hour lecture for $500 plus travel and accommodations, or a 2-day workshop at $1500, also plus expenses. I am dismayed that the TESOL organization has sometimes given unwitting support to what I feel is a blatant commercialization of an unproven method, and I am almost tempted to offer my services to travel anywhere around the globe to deliver a reasoned refutation of such pedagogical panaceas at one half the price, plus, of course, travel and accommodations! Let me share with you in the time that is left why I believe suggestopedy remains over-publicized and unvalidated.

Lozanov's book, which introduces his theory and presents what he believes to be the empirical evidence for its effectiveness when applied to language teaching, is no lightweight monograph. Interspersed among its some 377 pages are 47 tables, 53 figures- one of which includes an excerpt in Italian from Dante's *Inferno* as a sample lesson plan (forgive the pun, but what a hell of a class that must be to teach!), and a superficially impressive bibliography from the literature of neuropsychology and psychiatry of over 600 English items and over 400 items in Bulgarian. What emerges from a casual glance at the text, therefore, is a weighty scientific document, replete with references to empirical experiments. In fact, the entire manuscript is (and I choose the word carefully) *suggestive* of the notion that Lozanov speaks with the authority and impact of scientific proof. Unfortunately, a careful reading of this major work reveals that there is precious little in suggestology that is scientific and that viewed in the context of the enthusiastic claims about its success launched by Lozanov and his supporters, suggestology, at least as it is introduced by its founder, represents nothing more than an oversold package of pseudoscientific bunk! There are several reasons why I am led to be so sharp in my criticism; let me simply focus on one of them: Lozanov's continual propensity to dress both theoretical assumptions and practical claims with the superficial trappings of scientific experimentation.

This is not the appropriate forum to dissect Lozanov's experimental methodology. For a more detailed review of Lozanov's work, see Scovel (1979). Suffice it to say, beginning with the very first experiment reported in his manuscript, the serious reader will find that time and time again, Lozanov's experiments are sloppily constructed and shoddily reported, leaving the impression that all the references, tables, and figures are constructed to present a shiny veneer of empirical proof, while in reality, they appear at least, to me, to be singularly unimpressive. Lozanov never once bothers to describe the curriculum studied by his control subjects; he never once takes the trouble to detail the nature of the tests which reveal such constant and impressive memory retention scores in favor of his suggestopedic students; and he consistently fails to explain why the numbers of his subjects vary so dramatically from experiment to experiment—in one series of investigation alone, for example, there are as many as 416 and as few as 7, the scores improving as the numbers decline.

Many of Lozanov's experiments fail to eliminate or control intervening variables which, in all plausibility, may account for his test results much more effectively than the impact of the suggestopedic method. In corroborating the claim that suggestopedically memorized material is retained for long periods of time, Lozanov reports on the results of vocabulary tests which have been given as long as 22 months after the initial learning period, again with impressive success. What one does not know, unfortunately, is whether the words tested were part of the common foreign language vocabulary that the students had been employing both formally and informally during the period of time after they had first learnt these words. If they were tested under the same experimental procedures, I would display an amazing 100% retention rate of all the Thai words I even learnt in my first few weeks of Thai lessons fully 15 years ago, not because of the efficacy of the quasi-audiolingual method I was subjected to in a hot, humid, and noisy classroom by a semi-conscious teacher to a particularly dull student, but because of the simple reason that I have heard and spoken those words over and over again during the intervening period of time. The problem with Lozanov's experimentation is not simply that he employs a different experiment paradigm, one that is based on eastern European and Soviet philosophy and psychology. This particular point is raised in his defense by Jane Bancroft of the University of Toronto, who impresses me as the most prudent and reasonable supporter of suggestopedy that I have yet read (Bancroft, 1978). This defense rapidly evaporates however, if one compares his work to the father of Soviet psychology, Pavlov, whose early experiments in classical conditioning are more carefully constructed and more impressively reported than Lozanov's disappointing attempts at scientific validation.

There is also the important matter of the difference between memorization and learning. I do not need to belabor this difference to you, my fellow teachers in the audience, although, I hasten to add, it is still a source of pedagogical confusion in contemporary China where most teachers still believe that the seeds of memory will ensure a rich harvest of learning. Despite the one or two references to the possibility that language learning comprises more than memorization of lexical items and phrases, the entire thrust of Lozanov's pedagogical method is directed at enabling the student to memorize large quantities of material in short periods of time. Even in the solitary sentence of the book where the author attempts to extricate himself from the error of equating memorization with learning, he compromises his position in the very next breath. "The main aim of teaching," he writes, "is not memorization, but the understanding and creative solution of problems." "However," he continues, "the main obstacle encountered in teaching is memorization, automation, and the assimilation of the material presented" (Lozanov 1978:251). I find it impossible to distinguish any substantive difference between "main aim" and "main obstacle" in the preceding quote. More importantly, the innumerable references to experiments on memorization and the recurrent discussions of hypermnesia, *to the total exclusion* of references to "understanding" or "creative solutions" convince me that suggestopedy is an attempt to enhance memory and is not devoted to the far more ambitious and important business of language acquisition.

If Lozanov's enthusiastic claims about the efficacy of his method continue to linger only as unsubstantiated hyperbole, and if his principal interests remain focussed on hypermnesia and not on issues much more central to language learning, what are we left with as far as any possible new contribution from suggestopedy to our field? As far as I can see, very little. About the only alternative open to us is to follow that lead of some of the North American proponents of suggestopedy and borrow some of Lozanov's more worthwhile techniques, integrating them into an eclectic method which itself is not based on Lozanov's theoretical approach. Such ideas as the concert session, the use of role plays, and the assignment of new names and identities to the students in the second language classroom are all useful activities and could, indeed have, been incorporated into successful programs for both children and adults. The inclusion of these procedures into a more eclectic method is not, however, tolerated by the founder of suggestopedy: Lozanov is unequivocal in his stance against any tampering with his method. "Any eclectic combination of suggestopedy with other methods," he sternly warns, "brings a risk of lower effectiveness and of fatigue in the students" (Lozanov 1978:333). In view of Lozanov's defensive position about accommodating suggestopedy with a broad-based language teaching program, I do not believe it is fair for those who employ only bits and pieces of Lozanov's approach to label their method of teaching with the same brand name as the original, untainted product.

What then can I leave each one of you which is new and instructive, apart from this critical and perhaps intemperate appraisal of what might have appeared to be a major innovation in language instruction? I can only offer the old but simple fact that there are no easy paths to follow along the roads that lead to greatest promise. Certainly, learning a new language is one of those journeys with a goal of inestimable value. The fact that suggestopedy does not provide any short cuts should not deter us or our students, however, from pursuing our common destination. As the Chinese so eloquently put it, "qian li zhi xing, shi yu zu xia": "a journey of 1,000 miles begins with the first step." Let us take each step carefully, but with confidence. Godspeed to each of you on that exciting journey!

Methods Panel

Exploring Community Language Learning

Barry P. Taylor
University of Pennsylvania

Curran's Counseling-Learning/Community Language Learning (C-L/CLL) model (1961, 1972, 1976) is seriously misunderstood. It should be made clear from the outset that Curran's writings do not suggest that we revise or adapt our approach to education; rather, they advocate the adoption of a fundamentally different framework, one that is difficult to perceive if we approach it with a "questioning" rather than "questing" attitude (1972).

Understanding Curran's approach requires an openness to the totality of his perspective (Rardin, personal communication). One who cannot view it in this way may well feel, as I did when I began probing into the core of C-L/CLL, that it is a cultish, closed society. I found myself put off by its apparent lack of acknowledged limitations, its claims of success in the absence of rigorous statistical validation, and its lack of attention to the findings of linguistic and language acquisition research.

These criticisms may well be valid from our rigorously scientific point of view, but before we judge, we should try to catch the essence of Curran's relatively new, humanistically-based perspective. Once we have experienced it on its own terms, we will be in a much better position to evaluate it fairly.

I. A brief overview

1. *Counseling-Learning.* The theoretical underpinnings of the C-L model are complex. In brief, Curran maintains that learners approach education with personal learning conflicts, hostilities, and anxieties which can block intellectual learning. By incorporating the "whole person" of the learner into the educational process through an application of techniques gleaned from counseling psychology, especially "client-centered therapy" (Rogers 1965), the C-L model offers a means by which teachers and learners can deal productively with these feelings (Rardin 1976).

The C-L model places heavy emphasis on the power of group psychological forces. LaForge (1971:48-49) suggested that, for most traditional classroom situations, "the assumption seems to be that learning is an individual affair somewhat accidentally taking place in a group situation." In group learning, however, LaForge suggests, emotional support and acceptance are supplied by the group to each student. Individual defensiveness is reduced in this less competitive, less rejective, more secure atmosphere.

The result of Curran's research in utilizing counseling strategies to deal with the anxiety and fear experienced by many adult language learners led to the formulation of CLL.

2. *Community Language Learning.* It has been claimed that traditional language teaching approaches often produced students who had a good command of the "rules" of the target language, but were insecure and unable to communicate (Curran 1976). CLL's emphasis on making the target language immediately operational is, to a large extent, a reaction against these older approaches.

Today, however, language classes are fundamentally different from those of the past. The current trend toward learner-centered, contextually-real activities designed to involve the learners actively is paralled to many of the techniques used in the CLL class. But one who only notices the surface characteristics has missed the essence of CLL because its core does not reside in the activities or techniques themselves.

CLL is not a *method* of language teaching—it is an approach (Rardin, personal communication). As such, its focus is not on prescribing specific language teaching activities but on creating an entirely new approach to education—one in which the teacher gives up his sense of power and authority in the classroom in favor of entering into the world of the students. The teacher's role is to facilitate a feeling of security and self-esteem within his students which will enable them to rely confidently on him and each other, without feeling foolish in their ignorance, as they progress from a state of linguistic dependence to independence. The emphasis is on the intensely personal growth of each individual in the group as he enters into a "creative affiliation" (Rardin, personal communication) with the teacher. This approach requires a fundamental change in the teacher's perspective as he gives up his role of "answer man—knower of all," and adopts the counseling skills and sensitivities which will allow him to relate to the students from their perspective and become an "understanding counselor" who fully appreciates their struggles.

My CLL experience in Japanese illustrated many of these precepts. Because the class was not "teacher-centered," we took the responsibility to initiate much of our own learning. The result was that we were never pressured; as learners, we decided when it was time to move ahead or when we needed "space." One of the most striking observations from this language experience was how much Japanese we learned in only ten hours. The momentum, once started, progressed at an incredibly rapid pace. Perhaps the most surprising observation of all was the extent to which this class of strangers gave up some of themselves to support each other, even the slowest student, and create a sense of community.

From these experiences, I feel that this approach certainly has value and much to offer the language teaching profession. But, to quote Brown (1977:

371), ". . . we would do well to attend to the possible tempering factors . . . ," which I would now like to address.

2. Questing

Over the past few years there has been an abundance of research to suggest the enormous and variable role affect plays in second language acquisition (Brown 1973). Scovel (1978) recently discussed the difficulties entailed in defining "affective variables," and opted in favor of the broader term "learner variables," of which affect is just one. Learner variables range from physical to emotional to perceptual to cognitive differences among learners. Anxiety, which Scovel categorizes as just one affective variable among many, is one of the major targets of the C-L/CLL approach.

Like most learner variables, anxiety affects different learners in different ways. In his article, Scovel cites research to indicate that sometimes even anxiety can be facilitating rather than debilitating. Given the wide range of possible differences among learners, we must be wary about ascribing too much importance to any one variable or set of variables within the context of a teaching approach; it is unrealistic to assume that one approach will work equally well with all learners (Scovel 1978, Brown 1977).

While it is generally recognized that students learn better when the teacher is warm and supportive (Hanchey 1976, Scovel 1978, Stevick 1973), we must also realize that students can sometimes create or carry their own learning blocks, regardless of what we do. In a report evaluating the effectiveness of the CLL approach in training Peace Corps volunteers in the target-country setting, Rardin (1975) discovered that the combined variables of time pressure, difficulties in adjusting to the foreign culture, individual learning styles, and levels of motivation and expectation created a resistance to the CLL approach on the part of some learners. These trainees, rather than seeing CLL as a relaxing way to learn a language, found themselves more comfortable with the more traditional methods.

There are also mixed reactions to the student-directed aspects of CLL. Since the focus is on student-initiated conversations and gaming, linguistic forms are not presented in a predetermined order, but only as their need arises. Although most of the learning is inductive, occasionally deductive teaching is done by the teacher or one of the students (Rardin, personal communication). While the teacher is free to suggest or recommend any materials that he feels are appropriate, the students are given the responsibility to become personally involved so as to foster the self-investment which is presumed to maintain interest and motivation (Curran 1976).

Although this personal freedom was one of the most exhilarating aspects of my CLL Japanese experience, we must realize that my class was composed entirely of language teachers and, as such, we were well-equipped to direct our own learning. One wonders how classes with different kinds of students would proceed. Would the classes be like those Brown (1977:370) described when he wrote, "the initial stages of this process are grueling—there are days and weeks of struggle and sometimes intense confusion until some insights begin to be drawn by the learners"? How long would these students stay interested? How long would the faster learners encourage and support the slower learners and still feel that they themselves were learning?

Although our group entered the CLL experience fully recognizing that it was only an extended demonstration, students under personal or external pressures to succeed seem to react differently. The resistance of some of the Peace Corps trainees referred to earlier extended to the non-directed aspects of CLL as well. An objective report by a Peace Corps language specialist (Hanchey 1976:3) indicates that some of the new trainees found the approach frustrating. They issued "a plea for a return to the imposition of planned curricula and away from the 'freedom' of the new approaches."

And what of foreign ESL students in the United States? Are they more like my fellow CLL participants or like the Peace Corps trainees? My guess is that these ESL students would suffer from the same anxieties and pressures as the Peace Corps volunteers. The pressure to "pass" TOEFL, to be accepted at a college or university, to adjust to a different culture, not to mention a different language, and all in a short period of time, could contribute to impatience and quite possibly anxiety, hostility, or even culture shock. As indicated earlier, students carrying this kind of internal anxiety appear to be resistant to the CLL approach. Could it be that these learners, obviously instrumentally motivated, find the integratively motivated CLL approach incompatible with their perceived needs?

Leaving these considerations aside, is the CLL approach mechanically applicable for ESL classes composed of students with heterogeneous linguistic backgrounds? One of the ways that CLL builds security is by using the students' native language in class, at least at the early stages (Curran 1972, 1976). Although it has been pointed out (Rardin, personal communication) that a single monolingual teacher can utilize similar techniques, using only the target language, in a multi-lingual class, I have never actually seen such a demonstration. It would appear that even if such a technique could function effectively, the security of using what is known and familiar (i.e., the native language) would be lost. The period of reflection, which is an integral component of the CLL approach in that it offers students the opportunity to express their feelings, reactions, and concerns (LaForge 1971:55), would also lose its impact, especially at the lower levels where security is so important, if the students were unable to rely on their native language(s). The alternative would be the prohibitively expensive solution of providing native speaker counselors for each language represented in the class (Brown 1977).

And what of the language teachers themselves? As Brown (1977) has noted, CLL requires teachers well trained not only in linguistics, so that they could work comfortably without a text, but also in translation and counseling skills. In addition, these teachers would need to be native-like, if not native, in the target language as well as highly proficient in the students' native language(s). And where would these teachers receive their training, since most M.A. programs do not offer full courses of study in C-L/CLL?

These questions are not intended to detract from the enormously valuable contribution which C-L/CLL has made to the language teaching profession. While it may not be a panacea for all of our problems, C-L has dealt straightforwardly with many of the central issues facing us. It has questioned the role which we play in our classrooms and has proposed revolutionary changes which are well worth considering. It has made us aware once again of our students as people—those individuals who so often seem to be forgotten in our concern with new techniques and statistical research findings. It has taken some of the research in learner variables and made it operational in the classroom. But perhaps most importantly, it has forced us to look at ourselves and evaluate what we do. This is a challenge that must be taken seriously.

Part II

Curriculum Design and Implementation

Evaluating, Adapting and Innovating Language Teaching Materials

Christopher N. Candlin and Michael P. Breen

University of Lancaster

This article has two aims: first, to offer a comprehensive set of questions which can be used in teachers' evaluations of the materials with which they work; second, to offer a number of proposals which may be used as a basis both for the adaptation of those materials and for the design of new materials. The proposals for adaptation and innovation are based upon the growing recognition that teaching materials ought to confront the learner with examples of and opportunities for language use. The proposals are therefore intended to serve those teachers who wish to enable their learners to learn to communicate.[1]

In this article the term *materials* refers to any published or unpublished data in any medium or collection of media used for the purpose of language teaching and learning. We intend that the guidelines for the evaluation of materials can apply to single items as well as sets of materials.

The guidelines for evaluation and the proposals for adaptation and innovation focus on four main issues: 1) In what ways should materials be appropriate to the teaching situation and to the curriculum in which they are used? 2) What should be the content of the materials? 3) What teaching-learning activities should the materials promote? 4) What new directions are available for materials design? Each section of this article takes one of these issues and discusses a range of questions which arise.

I. Putting Materials in their Place

Language teaching materials need to be seen in their proper context. First, we need to know to what extent materials are sensitive to the teaching situation in which they are to be used. Second, we need to be sure that resource materials are in fact appropriate to the particular language teaching curriculum. We can specify each of these requirements more precisely.

1.1 Are the materials sensitive to situation? Many current materials seem to give the impression of ideal usefulness and applicability for any learner—at some assumed level of proficiency—in any place in the world. Such materials are not so much ideal as idealised; they do not take account of situation. Unless they are adapted to specific situations in some significant way they cease to be of any direct value to anyone anywhere. Such global applicability almost always involves compromise when the materials are actually used, compromise which attempts to translate the ideal into the actual situation. In fact, there can be no such thing as both ideal and universally applicable language materials. Even when materials designers have a particular situation in mind, they must recognize that their materials will always undergo processes of reinterpretation by the different users of the materials as teachers, learners, and classrooms impose their specific meanings on the materials. The only alternative to this reinterpretation process is when the materials come to prescribe the teaching and learning. Teachers can be driven by the materials, so the efforts of the learners to learn and use the language is subordinated to the specific requirements and conventions of the materials. In such cases, learners are obliged to suspend their own sense of reality and teachers are obliged to instill interest and involvement in something over which neither teacher nor learner has much possibility of control.

---

[1] The actual criteria on which the proposals are based derive from a definition of language as communication and a view of language learning as a communicative process. A detailed consideration of these criteria and their theoretical and research background is given in M. P. Breen and C. N. Candlin (forthcoming).

In what ways should materials be sensitive to situation? We are aware, first, that every language teaching situation reflects certain socio-culturally determined educational aims and philosophies; language teaching is supposed to serve the individual within the context of the state. We know, too, that any language teaching situation has its own specific human resources, its particular teachers and learners. Third, it exhibits its own material resources: its own financing, its facilities, and the time invested in its teaching and learning. Each of these characteristic constraints imposes its own demands on teaching and learning and, in particular, on the materials to be used.

To what extent as far as the matter of human resources are the materials appropriate to the teachers and the learners? We must remember materials producers are in the business of making materials attractive to teachers, not to learners, so materials producers might suspect that teachers will choose materials which match their own experiences and preconceptions of language learning. Materials must therefore look familiar. To what extent do the materials constrain rather than exploit the teacher's own competence as a teacher? Do they reflect or contradict the teacher's theories or beliefs about teaching and learning? Do they match the teacher's preferred methodology and open up new directions for it? Does the view of language implicit in the materials conflict with or complement that of the teacher?

Have the materials been learner-tested? Have they been evaluated by those learners who use them? Are there clear guidelines for learners on how the materials are to be used? If the materials have been learner-tested, and if they guide the learner in their use, which group of learners were the materials planned for and evaluated by? Often enough, learners are made to match materials rather than the reverse.

What other resources are assumed by the materials? Do they call for special classrooms and special timetables? Do they require media resources beyond the capacity of the institution to provide? What is the relative cost of the materials compared to its actual cost to the student? What is their cost relative to their life in cost-benefit and survival terms? There may be greater benefit in smaller quantities of varied materials than in classroom sets of one particular course.

Of course, materials share with all the other resources in language teaching the potential to overcome obstacles, but to do so their user must at least begin with a sensitivity to that situation. We have seen that materials themselves can become a constraint. Materials pose two basic questions: Is it the teaching-learning situation which needs to be changed in order to accommodate the materials? Or do the materials need to be replaced by others which will more adequately serve the existing situation?

1.2 Are the materials appropriate to the particular curriculum? Any language teaching curriculum answers three interrelated questions: What are the teaching-learning objectives of the course? How are these objectives to be achieved? To what extent have the objectives been achieved and have they been appropriate? So, a curriculum can be defined as an ongoing relationship between some specified purposes, the methodology of which serves as the means towards those purposes, and the evaluation procedures which inform us on the effectiveness of the methodology and the appropriateness of our original purposes. Materials need to be contextualised within this curriculum process; they should be seen as one resource within it and not the main driving-force of the curriculum itself. It is often the case, however, that materials will supply their own purposes, their own methodology, and they can provide criteria by which to assess learner's progress, judge our own methodology, and reconsider the suitability of our own original purposes. So, a harmony is needed between the hidden curriculum of the materials we use and the actual curriculum within which we work.

This article primarily explores the possibility of a positive, harmonious relationship between materials and the language teaching curriculum. First, though, we shall focus on the basic curriculum characteristics which can help materials to avoid the unrealistic ideal. Current interest in so-called functional language teaching offers an illustration of the problems of idealisation of materials.

If we examine purposes in such functional materials, we find first that the purposes they explicitly state are opposed to the purposes they actually imply. Second, the change in purposes they advocate is to be accomplished without parallel changes in methodology. The explicit purpose of these functional materials is to teach language as use by means of exposing learners to a sequence of language functions. Implicitly, however, the definition of such functions given in the materials is not a genuine view of language as use, where an essential characteristic is a variable and unpredictable relationship between function and form. In such functional materials, the functions are presented as if they could be static, separable and categorizable items, and are generally exemplified in some pre-selected sentences or prescripted dialogue. As such they lack the genuine contextualisation of real communication and have no appropriate function other than as models. Thus, two myths are perpetuated: first, that there can be some predictable relationship, out of context, between a function and those forms to which it is arbitrarily yoked. Second, that functions are an established and systematic framework upon which the teaching of form can be based. So, the teaching of functions is erroneously confused with the teaching of the formal or structural nature of language. This apparent change in purposes is a sham; such materials offer little or no change in methodology. Essentially the same methodology which once served the teaching and learning of language structure is assumed or proposed within many functional materials. Apart from tasks and exercises involving idealised data designed to give teachers and learners the impression they are dealing with language use—such as written dialogues and other unreal and often paradoxical texts—these materials imply that functions can be presented in some valid sequence, learned as separable and listable items based on consistent patterns, and can be memorized like the phrase-book phrases they most obviously resemble.[2]

Given the warning of this current example, what are the ways in which materials should be appropriate to the curriculum they serve? What are the questions that a teacher should ask of materials?

1.3 Are the materials appropriate to purposes? Here our first concern is with the specific target language, and secondly with the important characteristics of the learners who are taking the course.

1.4 Appropriate to Target Language? Is the language of the materials derived from sociolinguistic analysis of the target performance repertoire?[3]

This question implies that the data within the materials should be authentic to the target; they should come from native speakers and be intended for native speakers. We can also ask, however, what this target performance repertoire represents or is dependent on. Examples of target performance—what the native speaker would say in this or that event—are particular uses of the target community's conventions governing meanings or ideas, the conventions governing how these meanings should be shared and negotiated in interpersonal behaviour, and the conventions which govern language form, or the text, which realises such meanings and such behaviour. Target performance, therefore manifests the native speaker's knowledge of conventions which govern communication. Each example of target performance represents a meeting-point of these three knowledge systems: ideational knowledge, interpersonal knowledge, and textual knowledge.[4]

---

[2] A comparison of traditional structural-situational materials with more modern functional materials using the guidelines offered in sections 2 & 3 of this article will reveal these and other significant similarities. There are other misleading assertions and assumptions. One of these is that the *communicative* use of language equals a set of functions and that it is to be achieved primarily (and on occasion exclusively) through oral-aural practice. We have to deduce that communication does not take place between writers and readers.

[3] A performance repertoire refers to selective use of language which any curriculum takes as its purpose, however extensive and general or restricted and specific such a repertoire may be. Even native speakers control only a selection of varieties of a language, and such a repertoire will reflect the communicative performance experience and demands of that speaker. Language teaching can be seen to be concerned with such target repertoires in the sociolinguistic sense (Gumperz 1964) rather than with some global or ideal target language.

Examples of target performance also involve the communicative abilities of the users of the language. Communicating, either written or spoken, calls upon the ability to interpret, express, and negotiate meanings. The skills, either taken together or individually, depend upon these three crucial and underlying abilities. Reading, writing, listening and speaking are the means through which the abilities to interpret, express and negotiate act and become refined.

The distinction we have just drawn between the specific target performance repertoire and its underlying target competence leads us to replace the original question with the following:

1) Are the materials authentic to the target performance repertoire?
2) Are the materials authentic to the underlying knowledge and abilities which the target performance repertoire represents?

If authenticity to the target is a criterion we wish to impose on materials, then 1) would demand actual data from the specific target repertoire which is the purpose of the curriculum; 2) on the other hand, would demand any authentic data which themselves represent or draw upon those underlying knowledge systems and abilities which underlie the target in question, but which could in principle underlie a range of different target performance repertoires. Although different target repertoires of different curricula may well share common features, there is likely to be even more overlap between the target competencies which underlie them. So, the second question would imply that there need be no one-to-one relationship between the data in the materials and the actual target performance repertoire. Authenticity to target in terms of this question implies authenticity to the native speaker's communicative knowledge and abilities.

The distinction between (1) and (2) above leads to a third question:

3) To what extent do the materials need to be authentic to the target, in either sense of the term?

Materials can be seen as a means to the target rather than necessarily embodying the target itself. Can materials serve the process of teaching and learning rather than the product of teaching and learning? If so, then materials should serve as a link between learner and target and to do this they will have to draw on both the target and what the learner brings initially to the course, in order to exploit the potential relationships between the two. Here we require authenticity to both target and learner. Now the question becomes one of discovering what materials can draw on possible relationships of the two. They can either build on these possible relationships between the learner's initial repertoire and the target repertoire, or they can mediate between his initial competence—his communicative knowledge and abilities—and the target competence. Small overlap among competences will most likely be greater than that between repertoires, it makes sense for materials to be primarily concerned with developing the learner's underlying knowledge and abilities. In short, to focus on the learner's process competence, and be less directly devoted to target performance data. Here for us lies the possibility of distinguishing Process Materials from Product Materials.

---

[4] These terms derive from Halliday (1973). For further discussion of these knowledge systems and their implications for the language teaching curriculum, see M. P. Breen and C. N. Candlin (forthcoming).

We have, as a result, several criteria we can call on when considering how materials might be appropriate to target language: 1) authenticity to target performance repertoire; 2) authenticity to the knowledge and abilities underlying such a repertoire, authenticity to target competence; or 3) seeing materials as a means to the ultimate purposes, as serving the process of teaching and learning and representing a link between learner and target. These criteria are not necessarily exclusive: We may well demand that materials meet all three criteria to some degree. That degree would depend on answers to other kinds of questions we may ask of materials. One such question will be: in what ways should materials be appropriate to the learner who is entering the course?

1.5 Appropriate to Initial Learner Characteristics? It is an important truism that materials should go from learner to target. Whatever the purposes of the curriculum we need to consider therefore what the learner can initially contribute to those purposes and what will be the learner's own expectations of the curriculum. We do this for three reasons: first, to avoid idealisation of the curriculum and its materials, far removed from the actual learners; second, to exploit those potential contributions of learners which will be appropriate in various ways to the achievement of the target; third, to acount for likely differences between the material designer's view of the purposes and the individual learner's definitions. Such contributions and expectations will, of course, vary between learners and, significantly, vary within the learner as the learning progresses. Purposes of materials will need to accommodate this heterogeneity and variability in learner contributions and expectations if they are to link learner and target. In assessing the appropriateness of materials in relation to these learner characteristics, we can ask: 1) Is account taken of the learner's initial repertoire? (e.g. aspects of the mother tongue, aspects of the target repertoire the learner may imperfectly know etc.) To what extent? In what ways? 2) Is account taken of the learner's initial competence (initial communicative knowledge and abilities)? To what extent? In what ways? 3) Is account taken of relevant personal characteristics of the learner (e.g., age, sex, social and cultural identity etc.)? To what extent? In what ways? 4) Do the materials accommodate the learner's expectations of what language learning requires and the learner's expectations of what the target requires? (These expectations may stem from the learner's previous educational experience, and they are powerful factors in the learner's definition of the learning task) 5) Do the materials allow for the learner's own definition of learning needs? Are the learner's own specific interests and motivation accommodated? Not only are needs, interests and motivations distinct, they both vary among learners and over time as the learning progresses. How do the materials cater for this differentiation and variation?

In order, then, to reflect the particular purposes of the curriculum, materials must account for the specific target language and for specific learner characteristics. Target language can be viewed in terms of a performance repertoire or an underlying competence. Learner characteristics can be viewed in terms of particular contributions and expectations. Purposes in any curriculum, and the materials within it derive from these two main sources. Materials need to reflect both if they are to avoid idealisation. The role of Methodology is to draw these two sources together to enable the learner possessed of initial and changing characteristics to come to terms with the target.

1.6 Are the materials appropriate to Methodology? This is the central concern of this article. We have already argued that, in considering the question of authenticity to target, materials can be identified more directly with the teaching-learning process than with the product of teaching and learning. This argument importantly implies that the evaluation, adaptation and innovation of language teaching materials must be based primarily on methodological criteria. Our emphasis on the appropriateness of materials to initial learner characteristics further implies that materials should start from the learner and serve as a link between the learner and the target. Also, the idealisation of many materials stems from inappropriate purposes and a methodology which is insensitive to the process of teaching and learning since such materials frequently offer model target data (often idealised in itself) as the driving-force of the curriculum, requiring learners to act upon this product through imitation and practice. The process of teaching and learning, involving the development and refinement of the learning, is paradoxically absent from the materials themselves.

What methodological criteria within any curriculum should materials reflect? We need first to see the methodology of the curriculum in terms of two interrelated elements: content and process. These are both exploited by methodology as the dual means to the curriculum purposes. Content in methodology is the link between the content of the target language and that content which the learner brings to the learning. We can additionally see content as that which the teaching-learning process *works upon*. In this way, content is the servant of the process within methodology. Process is what teachers and learners undertake during language learning and it represents the language teaching curriculum in action in the classroom.

We can identify the methodological criteria for materials within a set of principles for the selection and organization of content on the one hand, and, on the other, in a set of principles for the design and selection of teaching-learning activities as representative of the process. Our point here is that materials will be appropriate to any language teaching curriculum to the extent that they serve as resource within the methodology of that curriculum. Materials which are primarily driven by purposes idealise themselves away from the language teaching process; as such, they cannot justifiably be seen as language teaching materials.

1.7 Are the materials appropriate to Evaluation? Every curriculum adopts its own Evaluation procedures. Materials can serve as a possible source of evaluative criteria within the curriculum, both by focussing on the evaluation of learner progress and on the curriculum itself, in terms of assessing the effectiveness of methodology and the appropriateness of the original purposes.[5]

Evaluation can be either summative (or end-of-course) or formative (ongoing and guiding), and materials can be exploited as a means of evaluation in both these roles. What is important is that the selected evaluation procedures of the particular curriculum must determine the ways in which, and the extent to which, the materials aid evaluation. The alternative to such assessment is to continue with evaluation-driven materials which are idealised from the process of teaching and learning in the same way as purpose-driven materials. Teachers are well enough aware how those materials which derive from externally standardised and imposed tests and examinations come to control and prescribe every aspect of the curriculum.

If we now ask the question of the ways and the extent teachers might wish materials to aid their evaluation procedures, we are led to consider several related problems. Are there criteria within the materials useful as a basis for evaluating learner performance? Are such criteria, present, sufficiently explicit to different learners so that the materials can act as a stimulus for learner self-evaluation and for evaluation by the learners of themselves as a group? This latter requirement would make materials contain criteria which can be differentially met by different learners. Materials in such a case would not assume any single or limited set of criteria for success, but allow for differential success. To what extent and in what ways do the materials offer feedback to the learner? Do they give clear guidelines on the direction learners should take after success or failure at a task? Are such guidelines and such feedback premised upon predetermined criteria, or do they allow for different learners achieving different tasks, in different ways and at different times? Should the materials provide the means for summative and/or formative evaluation? Can learners apply their own evaluation to the contents and activities of the materials? If we accept the significance of personal socio-cultural attitudes in language learning, and the importance of learners' values and emotions, then this becomes an important question for materials designers. We are aware that learning to communicate is a highly evaluative activity. Requiring learners to confront language use as if it was devoid of such affective characteristics is a form of idealisation. This is especially so if we wish learners to draw upon their initial knowledge and experience in language use, with its strong affective involvement. We could then demand of materials that they be open to learner evaluation by continually involving the learner's attitudes, values and emotions. Indeed, we might expect the materials to stimulate learner judgments about content and to provide activities which are characterized by strong attitudinal, evaluative and emotive potential. An outcome of such learner evaluation, of course, is an evaluation of the materials in terms of their own appropriateness to the learner's needs, interests and motivations in learning the language, and their effectiveness in helping the learner to learn. Since learners do evaluate the usefulness of materials against a range of socio-cultural and personal criteria, perhaps materials should exploit this exploitation rather than assuming that they can be value-free and infallible. Such exploitation would acknowledge the strongly affective character of language in use and language learning, and ought therefore to be seen as a strength rather than as an admission of limitations.

---

[5] The present article can be seen as an example of such curriculum evaluation in that it focusses on one resource which, as we have said, should mirror the whole curriculum.

2. Proposals

We have argued above that materials which are sensitive to the situation of teaching and learning, and which are appropriate to the particular language teaching curriculum, will be likely to avoid idealisation. We put forward proposals in this section directed against such idealisation and towards an appreciation of language learning and teaching as a communicative process.

2.1 Materials serve Methodology. Language teaching materials should primarily serve as a resource within methodology, and therefore the specific purposes or evaluation procedures of a curriculum are secondary concerns in materials design. Materials will help teachers and learners by incorporating content as a carrier of the teaching-learning process and by promoting activities and tasks which activate this process (see Section 1.6 above). Materials can act as a means to some specific purposes and as a means for evaluation, but they will do so for other than purely methodological reasons (e.g., they will contain content which derives from learner definitions of their longer term needs, interests and motivations). Materials design will be based on process criteria of teaching and learning rather than on product criteria (see Sections 2 and 3 which follow for a discussion of such criteria).

2.2 Authenticity is relative. There are three routes to authenticity in materials: 1) authentic data taken from the target performance repertoire (communication by native speakers for native speakers in the target community); 2) authentic to target competence in that it derives from the knowledge systems and communicative abilities underlying the particular target performance repertoire or a range of such repertoires; and 3) authentic to the process competencies of different learners during the learning-taking process competence to refer to underlying knowledge and abilities but those which the learner develops from initial competence during the process of teaching and learning.[4]

In materials design, therefore, we would propose: 1) that authenticity to the target repertoire is neither necessary nor sufficient for the teaching-learning process; 2) that authenticity to target competence may be necessary but not sufficient for the teaching-learning process; 3) that authenticity to different learners' process competencies is necessary but may not be sufficient for the teaching-learning process; and 4) that authenticity to both target competence and different learners' process competences is both necessary and sufficient. (Note that the involvement of target competence within materials is justified by the Purposes of the curriculum, not its Methodology.

---

[4] To these three we could add Widdowson's 'authentication' (Widdowson 1978) by which learners 'make authentic' subjectively data which are presented to them. Thus data can be authentic to a text-type (say a poem) but if it is not read as a poem, but as a source of grammatical information, it is not treated authentically.

2.3 Materials will allow for Differentiation. If we accept that authenticity to the learner within the teaching-learning process is a valid criterion in materials design, then differentiated materials are the necessary outcome. We would argue that differentiation needs to be accounted for in the following ways: 1) in terms of selection and organisation of content and the type and range of activities and tasks; 2) in terms of particular routes available to learners through both contents and activities; 3) in terms of particular pace or rates of learner progress; in terms of the particular media through which learners (as individuals or as members of sub-groups) act upon data; 5) in terms of the learner strategies (differential use of abilities) exploited variously by different learners even in the undertaking of common or shared tasks; and 6) in terms of the criteria applied by learners to the evaluation of their relative success or failure at particular tasks or undertakings (i.e., a range of possible formative criteria which learners may use to establish their own individual or group criteria for success or failure).

2.4 Materials can be a Product of the Process. Designers and teachers often see materials as a before-the-event necessity. Most often materials are seen as input for teaching and learning. Frequently this leads to unquestioned dependency on such materials, regardless of their effect. As an alternative, it is possible to see materials as emerging from teaching and learning, as ongoing achievements of the process. If we can accept the proposals above on authenticity and differentiation, then learner-created materials are a likely outcome. This outcome has another justification in terms of classroom management; materials created jointly by learners and teacher during the learning will act to reduce the obligation on the teacher continually to search for new and ready-made materials. Materials which are the product of one course or one activity within that course can be recycled later within that same course or within the process of another course. So, if we see materials as a resource within the teaching-learning process, we may also see them as a resource created by that process.[7]

3. What the Contents Page Might Not Reveal

We have been arguing that most language teaching materials emphasize content and take process for granted. Content has been seen as the substance of teaching and learning and in consequence methodology has acted as an inculcator of content, essentially as a device for force-feeding inventories of formal or functional items into the learner. Process was assumed by and subservient to content. As an alternative we have proposed that the methodology of a curriculum would not separate content from process and make the latter dependent on the former but, on the contrary, content would be made the servant of the teaching-learning process and acted on by that teaching and learning. Within methodology, content and process continue to be related and continue to help and influence each other, but content now adopts the role of carrier of the learner's process competence and the provider of data through which learners select and develop their own means towards the target competence.

---

[7] It is clear also that any process has within it a series of tangible (and in principle evaluatable) products, open to inspection by the learners and the teacher, which drive the process. It is, after all, by being sensitive to these that the process changes direction.

A chosen methodology in a curriculum can, as a result, provide criteria for two linked tasks. It can offer criteria within the set of principles for the selection and organization of content, the choice of which will be defined by the process being served. In addition, methodology can offer further criteria within a set of principles for determining the type and range of the teaching-learning activities which will work upon the content. It should be clear, however, that in most materials the principles which have been applied and the criteria used in their design are likely to be implicit rather than explicit, and in some cases, confused and inconsistent. In evaluating materials, therefore, we need to identify these implicit as well explicit criteria, and in adaptation and innovation we should strive to be consistent to the chosen criteria.

This section examines some principles underlying the selection organization of Content; Section 4 will go on to examine those underlying the activities promoted by materials.

3.1 What is the implicit theory of Language in the materials? Is the view of language that of a structured system of formal items—grammatical, lexical and phonological—much as in an analytical grammar? Is it a view which sees language in terms of functions and the uses to which these formal items are put, and therefore still a view of some finite set of items to be accumulatively learned? Is it, on the other hand, a view of language as communication, characterized by variable relationships between form and function: a view which sees language in terms of a dynamic process of the sharing and negotiating of meanings through the shared and recreated conventions governing particular communicative performances?

3.2 What is the Focus of the Content of the materials? What is the "content" of the content? We should distinguish here between the *explicit* focus and the *implicit* focus of the content. The author of the materials may adopt a particular theme or a series of specific situations as the explicit content, but the important implicit content may well be a sequence of categorized grammatical patterns or functional items.

3.3 How are the Contents Sequenced in the materials? What seem to be the criteria for sequencing adopted by the author both for the materials as a whole and for the units or sections within the materials? Does the author, for example, base sequencing on some serial progression from simplicity to complexity, or from frequency to infrequency? As an alternative to serial progression of either kind, the sequencing may be cyclic or modular. As such it may not necessarily depend on any linguistic continuum of simplicity to complexity or from frequency to infrequency. The sequencing may be carried by some thematic development, or by interrelated activities, or by a developing repertoire of interdependent skills. So, might the sequencing take the form of gradual expansion from a core; the learner's knowledge and abilities being simultaneously refined within frameworks which develop from a core of knowledge and abilities?

3.4 *How are the Contents Subdivided in the materials?* On what basis has the author broken down the content into units, and into parts and sections of units? What are the criteria on which chapters, units, sections, lessons exercises or parts of exercises subdivided? Typically, such subdivision may involve serialized structures, 'functions' or 'acts', and may be explicit or implicit. Explicitly the subdivision may involve episodes in some narrative, or a sequence of situations, while implicitly the subdivision may be by sentence structure or function.

3.5 *How is Continuity maintained in the materials?* How is the learner's progress guided from one part of the materials to the other? In what ways is a particular aspect of content earlier in the materials reinforced and/or refined in a subsequent part or parts of the materials? Is this reinforcement/refinement achieved by following some incremental, accumulative pattern of items, or an expansive, aggregative development which is retrospective and prospective? Is the continuity presented through formal items, or are other potential sources exploited such as, for example, the knowledge systems or skill use? Again, as with focus and subdivision, we may need to distinguish between explicit and implicit continuity offered to the learner.

3.6 *What direction should the learner follow through the materials?* Is the direction the user should follow made explicit, or is it once again implicit? More precisely, is the learner assumed to follow the direction of the materials in some predetermined way from the beginning, through the middle, and to the end? Alternatively, is the learner (and the teacher) able to take any of a variety of directions? Is the direction to be taken open to discussion and choice, or is it closed and prescribed, preventing entrance and exit at any chosen point?

3.7 *Proposals.* In the foregoing section we have set out some questions which can be asked in evaluating the content of materials. Here we offer a number of proposals concerning content which may guide its adaptation and innovation.

Theory of Language. We have already indicated that language as communication is characterized by the simultaneous realization of the ideational, interpersonal and textual knowledge systems. That, moreover, the social and interpersonal character of communication guarantees that it will be permeated with personal and socio-cultural values, attitudes and emotions; that, in consequence, learners of a language are involved in a process of sharing, negotiating and creating conventions against a background of previous and ongoing psychological and social experience. This view emphasises the interplay among the knowledge systems and the dynamic relationship between them in particular social events. In turn, this assumes that communication involves certain basic abilities of interpretation, expression and negotiation which act upon and are refined by the knowledge systems. The abilities are themselves realised through and refined by the surface skills of reading, writing, listening and speaking. Content, if it is to reflect this view of language as communication, should therefore act as a carrier of these related components which underlie communicative performance.

Focus. It follows from the above that the focus of content within a communicative methodology will be on knowledge which is both cognitively and affectively significant to the learner and which demands negotiation in an interpersonal context. Content will also focus on communication as a part of, and related to, other aspects of human behaviour.

Sequence. Accepting that the communicative process requires us to deal with variable conventions, we cannot assume that any step-by-step or accumulative sequence of content will necessarily be appropriate. The ordering of content will be affected by learners' alternative routes. Sequencing, as a result, derives from the state of the learner(s) rather than from any implicit logic of the content itself. Simplicity and complexity, frequency and infrequency, in any absolute sense, cease to have any value in a communicative methodology. As a result, sequencing must be a cyclic process where learners continually develop related aggregations of knowledge and refining the use of abilities, rather than accumulate separable blocks of static knowledge. Content becomes something to be moved into, out from, and returned to, by learners in a process of finer analysis and refined synthesis. Curriculum designers cannot, therefore, predict with any certainty the levels of content upon which learners will decide to evolve their own sequence in learning. They can only anticipate a richness of content which will activate the learners' process competences so that the ultimate target repertoire(s) become accessible to learners and the particular demands of such repertoires become recognized.

Subdivision. Given the communicative requirement to interrelate the components of the knowledge system—the ideational, the interpersonal and the textual—in interaction with the abilities involved in using this knowledge, subdivision in a communicative methodology cannot be based on serialized forms and functions. It has to be based on activities or tasks to be undertaken, where both knowledge and abilities are directed towards communicative performance. So, we are not concerned with units of content, but with units of activity which may guarantee this interaction within the knowledge systems, and between knowledge and the underlying abilities which work upon and derive from that knowledge.

Continuity. Within a communicative methodology, continuity does not lie within an ordering of formal linguistic items. It can be identified in at least four areas. First, there is continuity possible from one activity to another and from one task to another. Activities and tasks set up their own requirements for their progressive accomplishment and it is in the pursuit of these requirements that the learner finds tangible continuity. Second, continuity lies within communicative performance during the learning and teaching, both in terms of communication in lessons which will include sequences of negotiation, and within the discourse of language data in terms of coherent sequences of utterances. Third, continuity is provided through the ideational system. At the general level the learner has access to continuity of theme, while at the particular level the learner can have access to conceptual or notional continuity. Because ideational continuity is realized through a progressive refinement of textual knowledge—refining a concept, for example, implies a refinement of its linguistic expression and vice-versa—there is parallel continuity of meaning, grammar and discourse. Fourth, continuity resides in a skill-repertoire or cycle of skill-use during an activity or task. For example, the achievement of a particular communicative task might require a progression from reading to note-taking to speaking.

A communicative methodology would exploit these four areas of continuity as clusters of potential continuities for the learner, rather than expect one alone to be adequate. All of them can be inherent within any single task or activity. These alternative types of continuity offer two important advantages for the materials designer. They can serve the full process competences of the learners, and they can allow for differentiation. We need to enable learners to discover their own continuity, and hence establish the criteria for their own progress. In accomplishing some immediate tasks, learners will impose their own personal and interpersonal order and continuity upon the task, the communication which the task generates, the ideational and textual data which they act upon, and the skills they need to make use of in the achievement of the task. As a result, the progressive refinement of the learner's own process competence provides an overall learning continuity.

Direction. It follows from what has been said about sequence and continuity within a communicative methodology that content is not to be followed along some predetermined route. Nor is it the case that content exercises some external control and direction over the teaching-learning procedures. Choosing directions becomes part of the methodology itself, and involves activities of negotiation between learners and learners, learners and teachers, and learners and texts. Who or what directs content becomes a justification for communication about the selection and organization of content within methodology and about the various routes to be adopted by the learners through any agreed content. The direction of content, and the nature of the content itself, can be predicted within methodology only inasmuch as it serves the actual learning process of the participants in the group. Carrier content may be as diverse as the different Directions learners may take towards a common target performance.

4. Materials Working

As with content, we can identify a number of principles upon which to base activities, and, as before, there will be different criteria guiding the application of these principles. We should emphasise that our decisions earlier concerning content evaluation, adaptation and innovation will be closely related to decisions concerning activities and tasks. Activities work upon the content, and content serves the activities. It is to be expected that there will be symmetry between the criteria, and the decisions resulting from those criteria, both for content and activities in the materials. We should also take care not to identify the activities promoted by materials solely with those devices described as exercises. Every part of materials involves or implies an activity or a task on behalf of users. Thus, for example, reading a passage before answering questions about it is itself a task, as is listening to some audio data before working on them. Learners never passively react to data if their attention is directed towards them. In this sense there needs to be no real distinction between a piece of data and the tasks evolving from it.

4.1 What is the implicit theory of teaching/learning within the materials? In section 1.4 we made a distinction between product materials and process materials. We can make use of this distinction in considering the implicit theory or view of teaching/learning which determines the overall methodology of the materials. Those materials which are product-oriented assume that teaching is a matter of *transmitting* knowledge or information directly into the learner.[8]

Such a view is likely to assume that the knowledge itself will take a single path so that no real change in the knowledge occurs between the model data and the learner's own system. Learning becomes a receptive activity involving as little change as possible in the knowledge during its transmission and reception. Learners, in consequence, are all believed to learn the same thing at the same time in the same way.

Process-oriented materials, on the other hand, assume that teaching is a matter of cooperative negotiation with learners, involving knowledge and information in a process of joint interpretation and shared expression. Instead of transmission there is a concern for communication. Knowledge is never a static object "out there" but varies and changes because of its intersubjective nature. So, learning not only changes the learner's prior knowledge it also involves a continual change in the new knowledge. Learners in process-oriented materials are consequently assumed to be learning different things at different times and in different ways. However, learning need not be seen as entirely subjective. Differential learning is itself subject to the sharing and negotiation of knowledge typical of an individual learner seeking confirmation that she/he is learning and has learned, both through further negotiations with the data, and in negotiation with the teacher and/or fellow learners.

These alternatives of product and process are best seen as points on a continuum, representing theories within which there is variation.[9] Nonetheless we can ask which theory of teaching/learning do the materials assume or imply, and thereby discover at least which orientation they favour.

---

[8] For an interesting distinction between teaching seen as transmission and teaching seen as interpretation see Barnes (1976).

[9] Process materials can be seen as standing mid-way on a continuum between product-oriented materials and learner-oriented materials. For further discussion of such distinctions and such a continuum within a broader language-teaching curriculum framework, see M. P. Breen and C. N. Candlin (forthcoming).

4.2 What classroom Procedures do the materials assume? Do the materials predetermine what occurs in the classroom? What do they leave for the teacher and the learners to determine for themselves? Do the activities involved by the materials assume a similar or recurrent classroom procedure from lesson to lesson or do they assume variation and diversity in procedures? Should teachers and learners follow the procedures laid down in the materials or are they free to choose their own procedures and alter the materials accordingly?

4.3 What Participation do the materials imply? "Who does what?" in the materials? What opportunities are there for individual, small group and whole class involvement in activities? Does the involvement vary and if so, for what reason? Is the teacher assumed to be a participant or is she/he excluded from such participation and accorded some outsider or directing role?

4.4 What Roles are given to the teacher? Is the teacher a transmitter of knowledge, a model to be emulated in much the same way as data within the materials? Is the teacher an evaluator who shares the same criteria for success or failure as the materials? Is the teacher a guardian of how the learning ought to take place and be achieved? In brief, are the teacher's roles assumed parallel to those of the materials? Alternatively, is the teacher able to exploit the materials as a co-participant with the learners, adopting roles different from or complementary to those of the materials?

4.5 What Roles are given to the Learner? Are learners explicitly or implicitly assumed to see the data in materials as models to be imitated, practised, stored in memory and recalled at some later date? On the other hand, are learners expected to explore and solve unanswered problems within the materials, to interpret data, hypothesise and evaluate, and be given the opportunity to go beyond the information given? Is the learner expected to have little relevant knowledge, to be limited in his abilities and skills, or do the materials encourage the learner to exploit the relevant knowledge, abilities and skills she/he already has? Are tasks to be seen as difficult and of increasing difficulty as the materials progress, or is the learner's initial competence engaged from the start? In short, where do the materials place the responsibility for learning: with the learner? With the teacher? With the materials themselves?

4.6 How do the materials involve the Learner's Process Competence? We refer here specifically to the roles expected of the learners by the materials, from an examination of the activities and tasks within the materials. What scope do they offer for activating learner competence, whether initial, process or target?[10] Language learning, as in language use, depends on the simultaneous engagement of underlying knowledge and abilities. Nonetheless we can accept that some activities in the teaching and learning may focus on one or other of these components of competence. To do so is not to ignore the other components but to refine them simultaneously since, as part of a unity, they are inevitably involved. In undertaking any specific task, all the components of the learner's process competence will be activated to some degree or other. Thus, although the questions that follow focus on each of these components in turn, it is to be expected that materials will involve the learner's process competence as a unity at any point in the learning.

---

[10] See section 1.7 in this paper. We may regard these competences as equivalent to the distinction made between the learner's mother tongue, interlanguages and target language. For a more detailed discussion of the components of competence see M. P. Breen & C. N. Candlin (forthcoming).

How and to what extent do the materials involve the learner's knowledge systems? What ideational, interpersonal and textual knowledge must the learner exploit, activate and develop in working with the materials? What demands are made on ideas and concepts, knowledge of interpersonal behaviours, knowledge of language form? How are the learner's affects aroused? Is it an aim of the materials to engage the attitudes, values and emotional responses of the learner? Is she/he encouraged to evaluate and judge against a range of personal, interpersonal or social criteria? How and to what extent do the materials call upon the learner's abilities? In what ways do the materials activate interpretation and expression? Is negotiation within and among learners, and between learners and teacher, promoted and encouraged? How are learners required to deal with tasks? Are the abilities seen as interrelated or are they treated in isolation? Are they seen as underlying the surface skills or directly equivalent to skills? Which skills are exploited in the materials? Are listening, speaking, reading and writing to be used separately or is their potential overlap exploited? Are they treated in some externally motivated sequence, or exploited as a cycle or repertoire within some specific task, thus capitalizing on the developmental interrelationship between them?

4.7 Proposals. What criteria govern activities within communicative language learning materials? The following addresses this question by emphasising learner actions with and through materials.

Negotiating. Activities and tasks in materials should encourage negotiation: both personal or psychological negotiation between individual learners and texts or tasks, and the interpersonal or social negotiation between all participants in the teaching-learning process—between learners and teacher, and learners and learners. From a psychological viewpoint, negotiation is the process of interaction between that which is known in the learner's mind and that new knowledge which she/he confronts. From such negotiation both the original knowledge and the new knowledge are transformed and a synthesis accommodated in the learner's mind. From an interpersonal point of view, negotiation is a process of seeking and creating consensus between the ideas, affects and social identity of one participant and those of others within the communication. Throughout such negotiation, and as a result of it, ideas can be shared, affects recognised and social identities acknowledged. One important outcome of the process of negotiation is that the conventions upon which the negotiation will be initiated are themselves open to negotiation and recreation.[11] We can see, therefore, that both communication and learning—regardless of the knowledge to be learned—is a matter of negotiation.

Communicating and Metacommunicating. Activities and tasks within materials, as we have argued, should promote communicative performance. But such communicative performance should draw upon and extend learner's process competences, thereby engaging the developing knowledge systems and abilities underlying performance through the use of the learners' skills. Both teachers and materials need to distinguish between any target communicative performance and the learning of communication. Communicative performance in the classroom will be made up of both communicative and metacommunicative acts. The latter are those acts where learners jointly interpret, express and negotiate about learning and about communication. They can be of equal significance in the learner's development of process competence as the use and refinement of that competence through communicative acts. In consequence, materials need to provide for metacommunication as well as for communication.

Authenticating. In order to be authenticated by individual learners and groups of learners according to their own relevance criteria and their own developing process competences, activities and tasks within materials need to be differentiated. (see sections 2.2 and 2.3). Activities and tasks need to offer alternatives for who is to undertake what, when and how. They should be sufficiently diverse and rich to engage different learner's knowledge and affects, and abilities and skills, in interrelated and supportive ways. If they are called upon within the group, these differential contributions of the individual learners can be of benefit to all the members of the group. Such differential contributions to the group have dual potential: they can provide the basis for authentic negotiation and they can allow learners to take on a shared responsibility for the teaching-learning. Instead of presenting a problem, therefore, to the teacher, the diverse contributions of learners can be taken advantage of as sources for shared and agreed directions in the learning.

Co-participating. Activities and tasks in materials should involve teacher and learners as co-participants in the teaching-learning process. Teachers and learners have, of course, their own specific contributions to make towards this process, but teachers and learners may be given equal opportunity to negotiate, evaluate and provide feedback, and to offer new directions for the process itself. Activities and tasks need therefore be less concerned with providing answers than with posing problems for joint interpretation and with encouraging the sharing of expression as springboards for other activities and tasks beyond the materials. Simply, materials should be a provider of potential for activities and tasks as well as incorporating their own. Materials can, as a result, participate as a creative resource within a network of resources—both human and material.

---

[11] For interesting discussion of psychological and interpersonal negotiation see, for example, Neisser (1976) and Goffman (1972).

5. What New Directions Are Available in Materials Design?

4.1 Language Learning rather than Language alone and Process rather than Content. In the preceding sections we have emphasised communication as the overall purpose of the language teaching curriculum. Communication we have defined as a process of relating language forms and language behaviour in the context of social events. We have stressed that the conventions which link forms to behaviour are not fixed for all time, nor even certain among different participants in an event or across events but are variable and need to be constantly negotiated and accepted. Communication becomes a convention-creating rather than merely a convention-following activity, it is a social and interpersonal process. Learning to communicate is, as a result, not a matter of digesting a static and predictable body of knowledge, but learning how to interpret, express, and negotiate these conventions. As participants in this process, learners naturally start from their own communicative knowledge and abilities. Learning, for language learners, is the gradual development of communicative knowledge and abilities through the process of communication. To be authentic to this definition of the learner's task, materials will have a primary concern for the process of communication. Content in materials will, as a result, be redefined as any information and data which serves communication. Further, since learners will impose upon such information and data different interpretations, and will bring to them different contributions, authenticity to some particular target repertoire becomes a secondary concern in materials design. The activation of process competence in the learner, using whatever content as a carrier of this process competence—possibly including selections from or aspects of some target repertoire—will take priority in materials design decisions.

5.2 Activities rather than Models and Differentiation rather than Predictability. In fulfilling this concern with process competence, materials take on a new relationship with the learners and the teacher. Rather than directing the teaching and learning, they act as an interdependent resource between the learner and other learners and the teacher, and among all other available material resources. It becomes no longer necessary for particular materials to cover everything in the teaching and learning. They become an action-resource, serving as a means for sensitizing learners to the actual task of learning to communicate, a means for the observation and analysis of communication, and a means for the development of the knowledge and abilities which the learner contributes. As such, materials will be made up of units of activity with the overall objective of showing learners how to learn. Units of activity will seek to engage learner abilities—and their realisation through learner skills—and set these to work upon the themes and concepts of ideational knowledge in the context of interpersonal behaviour. To capture and maintain involvement of this process competence, activities are likely to be of high affective content, continually challenging the learner's attitudes, values and emotions. In materials as units of activity, texts, of whatever medium and type, become the servants and means of process competence and cease to be ends in themselves.

Units of activity within materials will be characterized as a wide variety of different problems and will incorporate tasks of different kinds. There will be no explicit continuity and sequence between units of activity; activities and tasks will be heterogeneous rather than homogeneous. Each activity will serve as a stimulus representing an initial framework for the learner's prerequisite knowledge and abilities which need to be brought to bear on some particular task or tasks. Materials as constellations of activities need no beginnings, middles, and ends in the usual sense. Of course, any activity will have a framework from which it plausibly begins, and it will lead to appropriate, through various, outcomes or products. However, if negotiation characterizes the teaching/learning process, and if different learner contributions are exploited within the activity, there may be no single predictable route through an activity nor through the materials themselves. This differentiation of routes in the learning can allow for the varied accomplishment of learner's individual and shared objectives.

5.3 Materials can pose Problems. Materials as an action-resource will be characterized by problem-solving. They will need to offer uncertainty, unanswered questions, and challenges to the curiosity of learners, rather than the accumulation of prescribed and unquestionable items of knowledge. In essence, the activities will involve the learners in the process of restoring equilibrium between their initial knowledge and abilities and that knowledge and those abilities incorporated within and demanded by the activity. The materials will be necessarily incomplete; they will encourage the learner to search beyond the information and data given. Activities as problem-solving would encourage the learner to seek help through the teacher's and other learners' interpretations and expressions, through other material resources in the classroom (dictionaries, grammars, and other manuals, for example), and through relating what is required of her/him in the real world beyond the classroom and to the appropriate available resources existing there. In this way, the activities within materials will serve as both starting-point and meeting-point for connections between the information and data within the materials and the learner's experienced world. Like the classroom itself, materials can function both as an observatory of human communicative behaviour and as a laboratory for the examination and analysis of such behaviour.

Designing problem-solving activities for materials is, of course, a problem-solving activity in itself. If we adopt units of activity in materials design, what kinds of design decisions are likely to confront us? Figure 1 generally illustrates the design process we might undertake in planning for an activity. Readers may perceive this process as exactly parallel to an activity they would actually promote in the classroom. The authors would suggest that the particular problem of activities design—and, therefore, materials design—could well be an appropriate problem-solving activity to be presented to learners during the teaching-learning process.

5.4 Materials design can begin with Learner Negotiation. In considering new

Figure 1. Decisions in Designing Activities Within Materials

Step 1—Input

1.1 What implicit objectives/solutions is the activity to promote?
Is there potential for different objectives/solutions?
1.2 What prerequisite learner contributions will be drawn upon?
1.3 What information/data might be appropriate to the activity?
In Content terms? (Section 3 of this article)
In Process terms? (Section 4 of this article)

Step 2—Formulation

2.1 What is the initial framework for the activity which relates each element of the Input—potential objectives/solutions, learner prerequisites, and appropriate information/data?
2.2 What is the nature of the problem(s) posed? Can learners recognize the problem and formulate or reformulate it?

Step 3—Search/Tasks

3.1 Are individual learner interpretations of the problem—and the information/data supplied—allowed for? Are individual learner solutions promoted?
3.2 How might the group or sub-groups undertake negotiation about individual learner interpretations as a means towards some group solution?
3.3 Might individual members of the group act as observers and evaluators—on the basis of the group's criteria perhaps—of (a) the solutions proposed and (b) the communicative process within the group? (i.e. individuals might act as task-evaluators and communication-evaluators).

Step 4—What Potential Outcomes? What Further Tasks?

Either 4.1(i) Sub-group expression of solutions proposed and sharing with other sub-groups.
(ii) Evaluative negotiation between groups as to the relative appropriateness of the various solutions—involving some agreed criteria.

Or 4.2(i) Group evaluative negotiation concerning the relative success/failure in (a) the overall activity/problem, (b) specific tasks, or (c) communication.
(ii) Group/sub-group design on an *instructional plan* for either (a) solving the type of problem(s) dealt with in this activity, or (b) developing the specific communicative knowledge and abilities required by this activity which the group or some of its members felt they had not sufficiently developed.
(iii) Peer teaching on the basis of such an instructional plan.
(iv) Recycle and Transfer in terms of *learner-designed* activities/problems which build upon and are generated by this particular activity/problem.

directions for materials design, we have so far suggested that materials can incorporate a view of language as communication, can directly reflect learning as a communicative process, and can contribute to this process through the provision of problem-posing activities. Each of these new directions imply that materials design may have, as its initial concern, the ways in which learners may interact with and through the materials. So, materials design could begin with the question: How might learners be encouraged to negotiate with and through the materials? Learner negotiation—as we have indicated in section 4.7—can be both a personal and interpersonal undertaking. Such negotiation interrelates with and builds upon the other communicative abilities of interpretation and expression. Accepting this interrelationship between communicative abilities, the materials designer can seek to provide answers to two further questions: 2) How might learners interpret the materials and develop interpretation through the materials? and 3) How might learners be encouraged to express and develop expression through the materials? Each of these questions present what is, perhaps, a new challenge to materials designers and teachers. We are obliged to uncover the complexities of the abilities of negotiation, interpretation, and expression in order to activate them in direct ways during the teaching-learning process.

We have previously suggested (section 1.4) that the four skills serve to realize these abilities and can act as a means to their refinement. However, the abilities underlie and act through all or any single skills, so the latter are the observable representations of the abilities in action. Studies of reading and writing, and speaking and listening, have uncovered a whole range of sub-skills appropriate to each main skill and which combine together in its use. It is very likely that each of the abilities of negotiation, interpretation, and expression are each similarly made up of appropriate combinations of sub-abilities. The sub-abilities of interpretation, for example, may be a cluster or complex of sub-abilities which continually interrelate with one another in order to help and refine actual interpretation. So, interpretation itself may be seen to be made up of the following kinds of sub-abilities: 1) Attention and initial discrimination; 2) Making sense (through global or general assimilation to particular analysis) and 3) Going beyond the given information/data (involving accommodation, inferencing, hypothesizing alternatives, evaluating and judging, and remembering).

We may similarly define the ability of expression in terms of sub-abilities of the following kinds: 1) Producing or representing (involving recall and recreation and the proposal of new meanings in new ways); and 2) Transferring or generalising (involving relating different information or ideas or relating the known to the unknown etc., transferring learning to other activities and tasks etc., applying new knowledge and refined abilities and skills, and evaluating the outcomes of expression or the feedback it produces). These are all possible examples of a number of sub-abilities which may combine and interrelate in the learner's efforts to interpret and/or express.[12]

In seeking to develop one or other of the learner's main communicative abilities, materials may activate directly one particular sub-ability or a related combination of sub-abilities. The recognition of the various sub-abilities within the main communicative abilities could provide the materials designer with both a richer and more precise focus of attention and a wider range of potential activities and tasks. One new direction available to materials designers, therefore, would be to begin from just these abilities and their sub-abilities. Since process-oriented materials emphasise the development of learner abilities—and their sub-abilities—we need to discover more precisely what they are. This discovery need not be separate from teaching and learning. It will emerge as a significant outcome from the teaching-learning process. In this way, materials design can be one of the ongoing products of classroom activities. This uncover-ing of abilities and this cycle of materials design, we have argued in this article, may most effectively emerge through a communicative methodology.

---

[12] Limitations of space have prohibited us from discussing and examining these sub-abilities in any detail. For a closer consideration of them and their place in materials design see in particular Part 2 of M. P. Breen & C. N. Candlin (forthcoming) and other sources referred to there.

Frequency: A Criterion for Syllabus Development

Joni Turone-Perkins

Southern Illinois University

In sum, ten grammatical structures were arbitrarily chosen for this study. Natural language data were gathered by taping a television talk show. Frequency counts were established for the ten structures by counting the number of times they were used. The frequency counts served the basis for a rank order of the ten structures. The frequency rank order was compared with the presentation order in three ESL grammar texts. There was no significant overlap between frequency and the orders of presentation found in the grammar texts.

1. Introduction

The EFL/ESL profession has experienced a profileration of teaching materials in the last five or six years. So enormous has been the volume of produced works that both experienced and inexperienced teachers are often dazzled by all of the paraphernalia which "guarantees" production of a student who is fluent in English.

But alas! How "expensively" disappointing and frustrating all of your efforts and time become when your best student responds to your question, "Where did you live in Iran?" with a perfectly grammatical, "two weeks ago." The WH-Question structure was taught only the day before!

Your immediate reaction leads you to re-examine the materials with which you have been working. A comparison of texts reveals similar exercises and explanations of usage and structure, but a different ordering of the grammatical structures in each set of materials. How did the authors arrive at their particular sequencing of structures? What criteria were used to order and even to select those particular structures?

2. Background

Such questions and observations led Ferguson (1966) to believe that many syllabi have been based upon impressionistic judgments and vaguely conceived theoretical principles. Charles and Agnes Fries (1961: 7) concur that "most of the sequences of the structured items, not only in the syllabus, but even in the textbooks are quite arbitrary with little or no justification."

Various morpheme acquisition studies dealing with frequency have led Dulay and Burt (1977: 118-119) to suggest that

> Frequency serves primarily to increase the probability that those structures which the learner is ready to process will occur, thus increasing the chances that the learner will be able to attend to and process them. If this is the case, one might expect that

> the rate at which acquisition takes place may be increased if maximal exposure of a precise sort at specific times over an appropriate period of time is provided.

It has also been pointed out by Larsen-Freeman (1974: 158) that

> . . . frequency of occurrence is some measure of the usefulness of a structure and furthermore, it is quantifiable and thus not left to the intuition of the individual author or teacher. In general, items that are in frequent use need to be taught before those that are more rare whether we are talking about formal items or grammatical categories.

In his book entitled, *The Threshold Level for Modern Language Learning in Schools*, J. A. van Ek (1977: 163) concludes that "we do not need to acquire grammatical patterns unless we intend to put them to immediate use. This may well be the criterion for selecting grammatical items."

Results in a number of experimental studies with vocabulary (Higa 1965: 172) determined that

> . . . familiarity is a significant variable in recognition, learning, and recall. The familiarity value of a word is measured by the frequency of its usage which is found in word counts such as *The Teacher's Word Book of 30,000 Words* by Thorndike and Lorge.

As Larsen-Freeman (1974: 159) appropriately asked, "If vocabulary lessons have been based on frequency occurrence, why shouldn't syntactic structures as well?"

In this paper, I will argue that frequency counts of grammatical constructions in natural language use should play an important role in determining the order of presentation of grammatical constructions in ESL grammar texts.

3. Methods and Materials

The questions to be addressed are:

1. What structures are more frequently used than others?
2. How does a rank ordering of grammatical structures compare with the orders of presentation in ESL grammar texts?

For purposes of this study, three popular ESL textbooks were used: *English Sentence Structure* by Robert Krohn, *Using English Your Second Language* by Dorothy Danielson and Rebecca Hayden, and Volumes One and Two of *Modern English: Exercises for Non-native Speakers* by Marcella Frank.

The justification for using these three textbooks in this study is that they are used in many different intensive language centers to present English sentence structure. Since the three textbooks were written with different purposes in mind, it would be unfair to the authors to expect them to present grammatical constructions in the same order.

To answer the aforementioned questions, I taped four hours of the *Phil Donahue Show*, which is a popular daytime television talk show featuring both well known and unknown people. The guests are informally interviewed by both Phil Donahue and the audience. Although I do not claim that the language sample obtained from these interviews is the epitome of natural language, this particular program was sampled for the following reasons:

1. A study such as this requires a large amount of data to make accurate inferences to general speech norms. It is extremely impractical to amass such large quantities of data on the street.
2. Due to what Labov (1972: 209) has termed the "observer's paradox," substantial quantities of truly natural speech are difficult to obtain under any circumstance.
3. Television talk shows, such as *Phil Donahue* do not use prepared scripts. Although a particular topic may have been selected in advance for discussion, the questions, responses, and comments are spontaneous.
4. The speech used is accompanied by much extralinguistic meaning, paralanguage, kinesics, etc. which augment natural language.
5. Previous research has determined that "conversation programs can give us a good cross-section of a population" (Labov 1972: 211).

Frequency counts were calculated for the following ten arbitrarily chosen grammatical structures:

1. 'to be' as a *notional* verb in simple present and simple past forms
2. present perfect
3. present progressive
4. yes/no question
5. passive
6. relative clause
7. noun clause (including wh-clauses)
8. modal
9. tag question
10. possessive marked by the inflectional 's on the noun

It should be noted that natural language use from a different source (i.e. a lecture or a soap opera) might have produced a different rank order based on frequency counts. For example, one would expect an interview show to contain many questions. Yet, the yes/no question did not have a high frequency of occurrence. If the data had been collected from a lecture or a soap opera, yes/no questions *might* have had a lower rank.

In order to avoid possible confusion, the structures counted adhered to the definitions found in the three textbooks being examined. For example, some people linguistically recognize the utterances, "right" and "huh", when marked with a rising intonation pattern at the end of a sentence, to be a type of tag. Because none of the texts deal with this phenomenon, tag questions were counted according to the standard textbook definition.

Interviews for the four particular shows included columnist and author, Ann Landers; comedian and actor, Steve Martin; politician, Edward Kennedy; and psychologists, Werner Erhart, Dr. Nathaniel Branden, and Dr. Will Schutz.

It should be noted that all speech (i.e. guest, audience, and Donahue) was analyzed.

4. Results

Table 1 contains a ranking from the highest frequency to the lowest frequency of the ten structures for each individual interview. It also lists a collapsed order of structures which will be called the final rank order of the four interviews.

TABLE 1

Correlation of Individual Interviews with Final Rank Order

Final rank order	Landers	Martin	Kennedy	Psychology group
to be	1	1	1	1
noun clause	3	2	3	2
modal	2	3	2	3
relative clause	4	5	4	4
present progressive	6	6	5	6
yes/no question	5	4	8	5
present perfect	7	7	7	7
passive	8	8	6	8
possessive ('s)	9	9	9	9
tag question	10	10	10	10
	rho = .98*	rho = .96*	rho = 94*	rho = .99*

* $p < .01$

The data in Table 1 indicate that there is considerable agreement among the ranks. Each interview rank order when correlated with the final rank order, produced a rho which was significant at the .01 level. Kennedy's rank order, more than any other single rank order, differed with the combined rank order: the rho between Kennedy and the combined rank order was .94 nonetheless.

To determine the degree of agreement or association among the ranks of the columns of Table 1, Kendall's Coefficient of Concordance was calculated and is referred to as Kendall's W. Kendall's W expresses the average agreement, on a scale from .00 to 1.00, among the ranks. For the four ranks in Table 1 the Kendall's W is .95. We can state, therefore, that the relation between the four sets of ranks is substantial.

Such a high degree of correlation reinforces the hypothesis that the four rankings demonstrate a high degree of similarity that would not occur by chance alone. The highest correlation obtained was between the psychologists and the final rank order. The strongest degree of correlation occurs with the group having the most interaction thus lending support to the final rank of structures.

The textbook order was based on the first formal presentation of the structure even though the structure may have been used throughout the entire text and/or divided between chapters in its presentation. 'To be' as a notional verb is not formally presented by Marcella Frank in *Modern English*. However, using a number ten rank for this structure cannot be justified since the verb 'to be' is used throughout the entire first volume. As you can see, it was issued a rank order of two since the use of this structure is *stressed* (although not taught) in a section of noun agreement with verbs.

Table 2 lists the Spearman rank order correlation for each textbook order with the final rank order. Krohn's textbook order approaches significance but does not achieve it. In order for there to be a significant correlation, the rho would have to equal or exceed .683. Neither Frank's nor Danielson and Hayden's textbooks shows a significant correlation.

TABLE 2

Correlation of Textbooks with Final Rank Order

Final rank order	Krohn	Frank	Danielson/Hayden
to be	1	2	1
noun clause	6	10	9
modal	4	8	7
relative clause	7	9	10
present progressive	3	3	5
yes/no question	2	6	2
present perfect	8	4	6
passive	9	5	4
possessive ('s)	5	1	8
tag question	10	7	3
	rho = .61	rho = −.27	rho = −.13

TABLE 3

Correlation of Individual Interviews with Textbooks

Interview	Krohn	Frank	Danielson/Hayden
Landers	rho = .65	rho = −.28	rho = −.05
Martin	rho = .66	rho = −.27	rho = .02
Kennedy	rho = .47	rho = −.22	rho = −.14
Group	rho = .62	rho = −.3	rho = −.08

TABLE 4

** Structure Frequency and Percent of Total

Structure	Frequency	% of total
to be	1393	36
noun clause	714	18
modal	671	17
relative clause	365	9
present progressive	313	8
yes/no question	279	7
present perfect	173	4
passive	144	3.6
possessive ('s)	26	.6
tag question	8	.2

** Total number of COMPLETE SENTENCES and UTTERANCES is 3895
Complete Sentences = 2183 or 56% Utterances = 1712 or 44%

A Spearman rank order correlation of each interview with each textbook yields a nearly identical correlation with that of the final rank order and individual textbooks.

It was not the intent of this paper to compare textbook presentations with actual structure usage in natural speech; however, I would like to note a few very obvious discrepencies between actual usage and textbook presentation.

Forty-four percent of all the discourse has been labeled as utterances (partial or incomplete sentences). Yet, none of the three texts contains one partial or incomplete sentence.

As mentioned before, the verb *to be* as a notional verb had the highest frequency of occurrence, yet, Frank does not formally teach this structure.

Thirty-seven percent of the yes/no questions were formed by a rising intonation at the end of a sentence (i.e., John is sick?). Such a figure represents over one-third of yes/no questions; yet, none of the three texts teaches this very basic question pattern. Frank (1972: 76) does include a short description of the Informal Omission of Auxiliary as in the example, *Need any money?* which has been derived from, *Do you need any money?* By ignoring thirty-seven percent of the responses, all three textbooks are doing a great disservice to their users.

The tag question appeared only eight times out of 3,895 complete sentences and utterances. This represents only .2 percent; however, an examination of the texts reveals that the three authors consider this structure to be a fairly important one. Danielson and Hayden have devoted their third chapter of approximately six pages to this structure while Frank and Krohn spend from two to three pages in formal presentation of tags. Had I included a type of tag by which a hearer confirms his statement with a rising intonation on the tag (i.e., *You're going,* $\left\{\begin{matrix}\textit{right}\\ \textit{huh}\end{matrix}\right\}$?), the frequency count for tags would have been much higher. As previously stated, this latter use of a tag was not recognized by any of the three texts.

Table 5 lists the three categories of noun clauses used in the interviews. The highest frequency occurs with the omission of *that* (52.3%). All three texts teach the optional use of *that* by placing it in parenthesis; none appear to emphasize the use of a noun clause without *that*.

Such a brief overview of usage is in no way intended to substitute for a more detailed and thorough examination of structure usage.

Discussion and Implications

I am aware that pedagogical implications are only assumptions at this point because there have never been any empirically based studies done to determine the effects of frequency structuring on the language learner.

Wilkens (1974: 13) points out that

> . . . since natural language environments do not involve careful limitation and control of linguistic forms and functions, adequate opportunity must be provided for

TABLE 5

Structure Frequency and Percent of Structure Total

Structure	Frequency	% of structure total
To be		
a. Simple Present	1232	88.3
b. Simple Past	163	11.7
Noun clause		
a. *That* Omitted	374	52.4
b. *That* Expressed	164	23
c. Wh-Clause	176	24.6
Modals		
a. *Can*	164	28.6
Could	28	
b. *Would*	173	25.8
c. *Will*	114	17
d. *Have To*	89	13.3
e. *Should*	38	5.7
f. *May*	19	2.8
g. *Must*	11	1.6
h. *Might*	10	1.5
i. *Ought to, Got to, Let's, Need to, Used to, Shall*	25	3.7
Yes/no questions		
a. Inversion	177	63
b. Intonation	102	37

> the learner both to experience the natural variety of language in use and to respond to the varied demands of normal language needs.

If one brings this idea into the classroom, and more specifically, to the textbooks one is using,

> . . . there remains a feeling that it is the real language which should be taught. This feeling appears to underlie adverse criticism of the English used in course-books on the grounds that English [American] people do not talk like that (Lee 1977:216).

Yet, "our knowledge of language in communication is too patchy for a thorough coverage of what would be needed" (Wilkens 1974: 11). Without large amounts of empirical data to establish criteria for materials design, selection, and usage, we will continue to grope in the dark, and to spend millions of dollars on ineffective materials.

Certain frequency is not the only criterion; furthermore, we must distinguish the two effects of frequency structuring. According to Johnson (1977: 7-8),

1. Infrequency of occurrence with reference to the student language needs may be a criterion for omitting a structure from a teaching programme, but once the decision to include a structure has been made, then supposed difficulty of acquisition seems more a valid criterion for deciding how much time should be spent on it.
2. Secondly, by using a graded grammatical progression the structural syllabus finds itself spending time on less frequent structures which delay the introduction of communicatively more important ones.

Until now, structural grading has been the most effective means of teaching language to the learner. This reasoning is valid; however, the criteria upon which we base this grading are faulty. Johnson (1977: 8) states that

> . . . justifications for a particular ordering which run 'we've taught the present, so now we'll move on to the past,' or 'we've done the indicative so now we'll do the imperative' are not justifications at all. They smack of a taxonomic linguistic description made without regard for communicative needs.

It would be very presumptuous to assume that the authors of any text, including the authors of the three textbooks studied today, did not give a great deal of thought to the ordering of their structures. Consequently, a clarification of the authors' reasons for ordering the presentation of the aforementioned syntactic structures would be immensely helpful.

This leads us to the last question: What practical application might this study have in the classroom? For purposes of brevity, I have listed the following:

1. Use of more natural language in the classroom by the teacher. Teachers should not monitor their own classroom speech in order to give the student a chance to hear "uncensored" language both in terms of grammatical errors and incomplete sentences.
2. Use of textbooks that have self-contained units as the three we examined today. This would allow one to use these books in any order.
3. Do not demand near native perfection from the student. We must realize that

> . . . using a language may be considered to be a performing skill: as with any other skill performing ability will vary greatly from individual to individual. It may be a waste of time to demand near native perfection from performers who will never be able to provide it. Our aim should be to ensure that the misuse of language is not so serious as to obscure communicative competence (van Ek 1977:164).

4. Encourage students to read, listen to the radio, watch television, and take part. Since we cannot teach them everything, we must be content to give them the basics for survival and the ability to understand in general terms what they hear and see. They will learn to teach themselves.

In summary, my study fits into a schema for education in the following way. Proficiency considerations are of utmost importance in setting up an instructional program. We ask ourselves how much of the target language has to be taught, what arrangement and order of presentation of material must be made and how does that order and presentation of grammatical structures, in this case, affect a student's learning them and control over them? And finally, how do we know that the student controls these structures? These are questions which go beyond the scope of this paper. However, the questions addressed in this paper are crucial to the curriculum design of an ESL program which concerns itself with the attained language proficiency of its clientele.

Developing a Format for Working on Language Use With Language Learners

Marilyn Martin-Jones

Stanford University

In the wake of the theoretical debate about the merits and demerits of notional syllabuses, the few accounts of materials development projects to appear recently can provide some timely insights into the feasibility of the proposals that have been made. This paper describes the thinking involved in an ongoing project aimed at developing a format for working on variation in language use in spoken English classes for intermediate-advanced EFL students at the college level. There is first a discussion of the difficulties involved in working with a strictly notional syllabus schema, then a detailed description of the type of syllabus that was eventually adopted: an "encounter syllabus" with an ethnographic dimension. The units of the syllabus are centered around a series of social encounters or speech events which form part of an overall narrative sequence. The teaching-learning activities built around these encounters are designed to reflect the interactive aspect of discourse and to involve students in an informal exercise in the ethnography of speaking, both inside and outside the classroom. This article describes these activities in some detail, illustrating the way in which they effect student and teacher roles.

1. Background

A great deal has been discussed about the whys and wherefores of developing language syllabuses which focus on communicative functions and language use. It is now generally agreed that there has been an overly narrow emphasis in language teaching materials on syntax, morphology and the referential functions of language. Innovations in course design, particularly at the intermediate and advanced level, are long overdue. Some interesting proposals have been made; Wilkins' proposal for a notional syllabus (Wilkins 1976) in particular has caused the most stir among teachers and course designers. Up to now, however, there have been relatively few accounts in the applied linguistics literature of the insights gained by those who have actually undertaken the task of designing syllabuses of this kind. From the point of view of both theory and practice, it is essential that accounts of the thinking that has gone into such materials development projects be made available, since it is only in a comparative light that the implications of the theoretical proposals for the actual design and format of language courses can be fully assessed.

The following is an account of some of the thinking that has gone into an ongoing materials development project at Stanford University, a project undertaken to provide a new kind of format for working on variation in spoken discourse with EFL students at the intermediate and advanced level. Given the current state of the art of syllabus design, there are only a limited number of alternatives open to the syllabus designer. I eventually opted for an "encounter syllabus" along the lines suggested by Candlin (1977), with a notional dimension incorporated into each unit. After a brief description of the syllabus design I will demonstrate how an ethnographic approach falls very naturally out of this kind of metacommunicative framework. As I developed this approach, both in designing the syllabus content and in working out teaching/learning activities in the classroom context, I found that it provided a particularly dynamic means of linking the spoken discourse and language activities in the classroom with the speech communities immediately beyond the classroom. The real advantage of an ethnographic approach is the opportunity it affords for engaging students in an active learning process, for involving them in informal but systematic exercises in the ethnography of speaking outside the classroom: making observations, documenting and comparing their observations and conducting surveys of language use. In this paper, I will describe in some detail the kind of teaching/learning activities that I incorporated into the syllabus. These are but a few examples of the kinds of activities that could go along with an ethnographic approach. I am convinced that there is still much more that language teachers and syllabus designers can learn from the work that ethnographers do. As Hymes has pointed out, the strength of the ethnographic approach is that:

> It is continuous with ordinary life: much of what we seek to find out in ethnography is knowledge that others already have. Our ability to learn ethnographically is an extension of what every human being must do, that is learn the meanings, norms, patterns of a way of life. (Hymes 1968:18)

2. Objectives

I had several objectives in mind in developing the materials:

2.1 Linguistic content of syllabus: To focus on language use by A) dealing with the communicative functions of language, with the way people get things done with words in day to day language transactions; B) by coming up with a syllabus design which would reflect, as far as possible, the interactive aspect of spoken discourse; and C) by getting away from the conventional pedagogical strategy of treating language as a homogeneous system: focusing instead on the stylistic variation associated with differing degrees of formality.

2.2 Teaching/learning activities: To focus on learning by doing: with listening and speaking as complementary activities by A) providing students with ample practice in interpretation of spoken discourse instead of giving them ready-made explanations; and B) by involving students in task-centered communicative activities inside and outside the classroom—to get them to *use* the language to learn more about it.

3. Selecting a Syllabus Schema

3.1 A notional syllabus? Wilkins has proposed that language syllabuses should be built around "general aspects of meaning and use" (Wilkins 1976:21) rather than around a set of grammatical structures or situations. He conceives of what he has called a "notional syllabus" as having: 1) semantico-grammatical categories dealing with, for example, ways of talking about time, space, motion, quantity, case relations; 2) categories of model meaning; and 3) categories of communicative function.

- Judgment and evaluation
  - Valuation
  - Verdiction
  - Approval
  - Disapproval
- Suasion
  - Inducement
  - Compulsion
  - Prediction
  - Tolerance
- Argument
  - Information
  - Agreement
  - Disagreement
  - Concession
- Rational equiry and exposition
- Personal emotions
  - Positive
  - Negative
- Emotion relations
  - Greeting
  - Sympathy
  - Gratitude
  - Flattery
  - Hostility

It is this third component of notional syllabuses that presents the most problems for linguists and course designers alike. Wilkins rightly points out that up till recently, language instruction has focused almost exclusively on the use of language to report and describe. Intuition alone tells us, however, that other communicative functions should be taken into account in the design of a language teaching syllabus. Language is also used to get things done, to express emotions, to build a rapport, to mark social distance, and so forth. Wilkins gives us a broad taxonomy of potential communicative functions and attempts to match these functions with a graded schema of syntactic forms which are potential surface realization of these functions, for example:

Information sought:	'Question':
—request, question, ask.	Information seeking is likely to be an important aspect of a learner's language use.

(a) Interrogatives
(b) Declaratives + question intonation
(c) Question-word questions

When
Where far
What How + much
Who ?
What (time)

'Request'
Would you shut the window, please
(Would you mind shutting . . . .)
(—Wilkins 1972)

Wilkins' overall taxonomy is largely based on Austin's inventory of speech acts (Austin 1962) which is essentially a list of performative verbs or surface structure realizations of speech acts compiled on the basis of introspection.

3.2 Problems involved in working with a notional syllabus schema. While Wilkins' proposal for a notional syllabus represents a bold step forward in the right direction, I feel it raises a number of unfathomable theoretical issues that make it unworkable in practice. There are bound to be problems involved in carrying over the abstract constructs of speech act theory to the practical task of syllabus design. Not the least of these problems is the danger of oversimplifying the relationship between language form and language function. Language forms can be multi-functional. Even a simple lexical item such as: *OK* can be construed by the hearer as 1) compliance with a request; 2) the granting of permission; 3) the expression of approval, and so forth. For the most part, function can only be understood with reference to the context of the discourse itself and to the communicative context; i.e., the relative status of the participants, their goals, background knowledge.

Candlin (1976) has shown how the assumption that there is a one-to-one mapping between speech acts and their surface forms has led to the adoption of a static "item-bank" approach to syllabus design. He gives the following example:

> The usual procedure is to cite (speech) acts with a performative label and to associate with them a number of different surface structure construction types...
> Suggestion: Why don't you.....?
> I suggest that.......
> We could...........
> Let's.......
> How about.....?
> Can't we.....? (Candlin 1976:249)

The important point here is that presenting lists of lexical items and grammatical structures under the rubric of some loosely defined communicative function provides the language learner with no more guidance than a tourist phrasebook constructed around an ad hoc series of situations. The function of an utterance which begins with *Why don't you. . . . . ?* can only be determined with reference to what went before in the discourse and to the appropriateness conditions for the production of the utterance. Thus, it can only be interpreted as a suggestion *if* a suggestion has previously been elicited, either verbally or non-verbally, by the interlocutor(s) who is/are the referent(s) of *you*, or, if the context is such that the speaker has the authority to volunteer an unelicited suggestion. These are the kinds of pragmatic rules that would have to be incorporated in a strictly notional syllabus if such a syllabus is to be devised!

The other set of rules that need to be incorporated are those governing stylistic variation. In the examples of surface realizations of the function suggestion given above, there is no indication of which constitutes a more or less formal alternative and in what kinds of communicative settings they might be appropriately used.

Another problem with a strictly notional syllabus is the question of the boundaries of speech acts and their surface realizations. A speech act is not merely realized as a simple lexical item or utterance. As Coulthard has pointed out, in Wilkins' blueprint for a notional syllabus, "there is no indication of the size of the communicative acts, whether they are clauses, utterances, exchanges or transactions" (1977:141).

Aside from these form-function issues, there is the practical problem of what and how many speech act types are to be included in a notional syllabus. Speech act theory offers no guidance on this score. Searle has implied that there may be some "basic illocutionary acts to which most or all of the others are reducible" (1969:69) but he offers few clues as to what these might be. The selection of speech acts for inclusion in a syllabus remains a practical problem for the syllabus designer.

Most of the theoretical problems involved in working with a notional syllabus schema stem from the fact that the pedagogical orientation is from a single abstract discourse function to its textual realizations. The rules governing the way in which a linguistic form or constellation of forms come to convey a given communicative function cannot be made pedagogically explicit; they are complex rules involving reference to the presuppositional structure of the discourse itself and to the communicative context as well as to the shared knowledge of the interlocutors.

Quite apart from the theoretical problems involved in working with a notional syllabus schema, it also seemed counter-intuitive to peg the unit of a syllabus onto an abstract discourse function. I decided instead to adopt a more inductive approach: building the units of the syllabus around samples of spoken discourse and emphasizing interpretation rather than explanation.

As I mentioned at the outset of the paper, this approach is closer to what Candlin has called an "encounter syllabus." For the students, the emphasis is on practice in interpreting spoken discourse encountered both inside and outside the classroom. For the instructor, the goal is not to present ready-made interpretations of spoken texts but to explore with the students alternative ways in which the same functions might be expressed. The overall purpose of the syllabus is to make students more aware of how the different aspects of speech events can govern what gets said and how it gets said. As Candlin puts it:

> Learners need what we might call *encounter syllabuses* where they are progressively sensitised to, say, Hymesian components of speech events so that they can themselves analyze and compare the realization of encounter types. (1977:xi)

4. The Syllabus Design

The syllabus is built around a series of verbal encounters or speech events which fall into a narrative sequence. Each unit is labelled as if it were a new episode in this sequence. The narrative is centered around two characters: a male and a female student who are both arriving for their first quarter at an American university. The narrative sequence follows the kinds of communicative contexts and the kinds of language transactions that a foreign student arriving in the United States for the first time would be likely to be involved in, as well as the kinds of topics he or she would be likely to talk about. In this way, the syllabus format is designed to reflect as closely as possible the potential language needs of a specific student population, language needs being broadly interpreted in Richterich's terms as stemming from an interaction of communicative functions with settings, topics and notions, as well as the actual grammatical and lexical structure of the language. (Richterich, 1973)

Narrative sequence

The narrative sequence unfolds as follows:

Unit 1. ARRIVING
  Asking for directions
  Giving instructions
Unit 2. MAKING CONTACT
  Greetings
  Conversational openings
Unit 3. MEETING NEW PEOPLE
  Introductions
  Self-introductions
Unit 4. THE END OF AN EVENING
  Saying goodbye
  Making arrangements
Unit 5. A PLACE TO LIVE AND A BIKE FOR SALE
  Asking for information
  Making requests
Unit 6. REGISTRATION DAY
  Asking for permission
Unit 7. AN APPOINTMENT WITH AN ADVISOR
  Advise
  Making plans etc. . . .

In its present form, the entire course has fifteen units: the length of the course being dictated primarily by practical pedagogical considerations rather than by any of the linguistic criteria discussed above. The narrative theme is a very loosely-organized one, so that the units do not have to be followed in a rigid sequence. Rather, they can be chosen according to the needs and the level of the students. The theme does, however, serve a number of useful purposes:

1) It generates a set of "texts" of spoken discourse for interpretive practice; 2) it provides thematic continuity and a means of linking the language activities based in the classroom to the environment and to the immediate experience of the students since the episodes portrayed in the dialogue represent the typical language transactions of daily life on a college campus; 3) the two characters

are portrayed in a wide variety of communicative encounters, both formal and informal which serves to highlight the stylistic variation which provides one of the main focuses of the course; and 4) the notion of formality is defined primarily with reference to the norms and expectations of the speech communities that foreign students are most likely to come into contact with.

4.1 Notional dimension. A notional dimension is built into each unit. This does not mean that the entire unit is built around an abstract discourse notion, but one or two speech act types naturally occur in the speech events that form the basis of each unit. Thus, for example, in Unit 1, there are three samples of spoken discourse in the listening component:

A. At the airport. One of the students has just arrived. She asks how to get to the campus. The airline representative gives her instructions as to how to get there by bus and then directs her to the bus stop.
Speech act sequence: Asking for instructions
Giving instructions
Giving spatial directions

B. Near the campus. The other student is arriving by car. He stops to ask directions to campus.
Speech act sequence: Asking for directions
Giving spatial directions

C. On campus. At the student union. The same student asks how he can get ahold of a student he knows who lives on campus. He is directed to the registrar's office (a complicated route involving going to the third floor of the administrative building by elevator) and is told who to ask for.
Speech act sequence: Asking for instructions
Giving instructions
Giving spatial directions

In this way the speech acts or notions are presented as part of a pattern of exchanges in the context of a conversational interaction, not as isolated utterances. These instances of discourse are designed to capture the kinds of linguistic routines that are an essential part of a native speaker's communicative competence. Verbal interactions on the telephone are typically routinized, thus a telephone sequence is built into both the listening component and the spoken component of each unit.

The motivation for incorporating a notional dimension is two fold:

1) Intuitively and empirically, speech acts represent more natural units of discourse; and 2) the lexical labels for speech act types, such as *suggestion, request, apology,* and so forth, form part of a folk taxonomy of communicative functions for English speakers. Since students coming to the task of learning English as a foreign language have similar, if not equivalent, taxonomies of communicative functions in their own linguistic repertoire, it seems only appropriate to draw on this tacit knowledge in some way in developing materials for teaching intermediate/advanced students.

4.2. Sociolinguistic aspects of speech events. At different junctures in the narrative sequences verbal encounters of differing levels of formality are introduced within the same unit. The aim is to illustrate how different aspects of speech events, such as the identity, status and number of participants, the channel, the setting, the topic, the motives and goals of the participants, effect the way in which people choose to talk to each other. These aspects of speech events have been schematized as follows by Hymes:

S — Setting
P — Participants
E — Ends (the motives and goals of participants)
A — Act sequences (message content or topic)
K — Key (tone, manner or spirit)
I — Instrumentalities ((a forms of speech: codes, registers, dialects, and (b) channel)
N — Norms (norms of interaction of interpretation)
G — Genre

(Hymes, 1972:65)

The aim of the syllabus is to develop students' awareness of how these factors influence the way that people use English in the kinds of speech communities they are likely to come in contact with. Broadly speaking, these communities have been defined as the American college/university campus and those social, bureaucratic and commercial networks typically associated with a campus: from the acquaintances, friends, administrative staff and professors a student would normally interact with on campus to the people a student would typically come into contact with in applying for a driver's license, in buying an airline ticket, in visiting a dentist or going around to garage sales.

These aspects of speech events are brought to the students' attention in the first lesson unit. In the student workbook, they are given the following set of questions, based on Hymes' schema. The questions appear on the first pages of the students' workbook and they are encouraged to add any other questions they consider to be relevant:

1. Where is this conversation taking place? And when? Does this affect the way that people speak? (SETTING)
2. How many people are taking part in the conversation? Are they the same age or sex? Do they have equal status? Or is one of the people more important than the other(s)? (PARTICIPANTS)
3. Why is the conversation taking place? What does each speaker want to express? Does he/she want the other(s) to do something? Is he/she expressing some feelings? Or is he/she planning or promising to do something? (ENDS)
4. What are these people talking about? Does this affect the way that they talk? (ACT SEQUENCES)
5. Is this a serious conversation? Or are the speakers joking? (KEY)
6. Is the conversation taking place on the phone? Are the participants speaking in a formal or an informal way? (INSTRUMENTALITIES)
7. Is this the usual way to express this kind of thing? How does this compare with the way it would be expressed in your language? (NORMS)

8. Is this a special kind of situation, e.g., a lecture, a debate, or if is just an ordinary conversation? (GENRE)

These questions are designed to guide students in their interpretation of the instances of spoken discourse they will be encountering both inside and outside the classroom, and, to help them judge appropriateness. They can be addressed to the class to stimulate a discussion, or to the instructor whose role would thus be essentially that of a native speaker informant. Outside class, the students ask the same questions to native speakers as part of a systematic exercise in participant observation and informal discourse analysis, which will be illustrated later in this paper.

4.3. Syllabus content. The syllabus is designed to be as flexible as possible so that it can be used with students at varying levels of proficiency and in different programs. Each unit is divided into four phases, then subdivided into core and optional components: a listening component, a spoken component and a writer follow-up. The sequencing of the phases is illustrated below with reference to Unit 3: MEETING NEW PEOPLE.

PHASE I: Focusing Phase

Listening/Spoken Component: Discussion of cross-cultural differences in verbal (and non-verbal) conventions for:

(A) Introducing other people

(B) Introducing oneself—When? to whom? how?

(C) Use of terms of address; addressing strangers, attracting people's attention, etc.

(Questions provided in teachers' notes and in student workbook.)

Written Follow-up: Students are asked to write a short paragraph on one of the above.

PHASE II: Intensive Practice

Listening Component: Students listen to short "texts" of spoken discourse.

Speech act sequence: Greetings
Introductions
Identification of discourse participants

The texts used here are scripted conversations which are designed so as to reflect spontaneous spoken discourses as closely as possible. The conversations are marked by false starts, long pauses, interruptions, the use of attention getters and other typical features of spontaneous spoken discourse. I opted to use scripted conversations to illustrate clear, contrastive instances of stylistic variation and to include specific discourse features within a short stretch of discourse (about two to three minutes). These proved to be impossible to capture in an equally short sample of spontaneous spoken discourse.

Spoken Component

There are several suggested activities here, which are appropriate for different levels of proficiency and for different types of classes.

A) Students ask and answer questions about the sociolinguistic aspects of the speech events.

B) Students pick out lexical and structural items that are marked for formality or informality. These are placed on a rough continuum of formality. (On the board, in the student workbook, etc.)

C) Guided practice with new lexical and structural items. The emphasis is on *authenticity* of language activity.

(i) *Communicative drills* that provide practice with discourse sequences: In this case, greetings—introductions—identification of discourse participants. Students work through the drills in pairs in class or in the language lab.

(ii) *Task-centered activities:* These are of two types:

a. Those centered around some non-verbal context, such as a map for practicing giving directions in a specific context: inside a building, for example. In Unit 3, this part of the spoken component involves a role-playing exercise centered around a map (of a city) and a telephone. The discourse sequence is as follows:

—self-introduction
—acknowledgement
—invitation
—spatial directions

b. Those centered around some aspect of language use. For example, in Unit III students are asked to write a formal letter introducing themselves to a professor who is working in their field. The main objective of the letter is to ask for copies of several papers that have been written by this professor. The letter opens, nevertheless, with a self-introduction and the students are asked to work in pairs on the opening paragraph. Then, in a similar arrangement, they are asked to write to a 'friend of a friend' here in the U.S. The sequence of the discourse is as follows:

—self-introduction
—announcement of impending visit to city where this person lives
—expression of desire to meet this person.

This letter is an informal one. Although this is ostensibly a written assignment, it is included in the spoken component because the students are talking about stylistic variation, and, simultaneously making suggestions to each other, accepting and rejecting those suggestions and so on.

Written Follow-up

Again, in this component, there are a number of different activities suggested—to be selected according to the students level of proficiency, the type of class, the amount of time available, the homework load and so forth.

A) *Student workbook:* The students answer questions about the speech events. The questions are designed to elicit answers containing the key lexical items and syntactic structures in the "texts" of spoken discourse; e.g., those marked for formality/informality, etc.

B) *Student workbook:* The students note down those key items that they intend to incorporate into their productive repertoire. They can be encouraged to use some useful mnemonic such as a flow chart, or stick figures and speech bubbles, one set for formal and one for informal interaction.

C) *Student workbook:* Students read transcripts of actual samples of spontaneous spoken discourse with whole utterances blanked out at certain junctures. They then fill in the blanks with what they consider to be the most appropriate utterance at that point in the discourse. This can be done in class in pairs, as an individual homework assignment, or alternatively students can be encouraged to discuss the assignment with a native speaker outside class.

PHASE III: Extensive Practice

Listening Component

The listening and interpretive activities are essentially the same as in Phase II. This is an optional listening phase designed for more advanced students and can be done entirely in the language lab with the help of the workbook.

The texts used in this phase are unscripted conversations. However, because of the practical problems involved in obtaining good recordings of the specific discourse features and speech act types that I had in mind from actual spontaneous speech exchanges, I opted instead to record members of the speech community role-playing typical conversational exchanges in a series of predetermined settings. This proved to be an effective way of getting at speakers' stereotyped representations of what the ground rules for speaking were in those settings. Stereotypes do tend to be somewhat exaggerated, but consistent. This approach to collecting samples of natural spoken discourse provides a means of eliciting examples of language transactions set within a well-defined context and with differing levels of formality.

Spoken Component

A) *Role-playing practice:* This is centered around a series of well-defined communicative contexts suggested in the materials for use in class. The role-playing activity can be structured in different ways; one fairly productive technique is to record the students as they are engaged in the role-playing and then to go back and discuss appropriateness of use.

B) *Telephone practice:* This exercise is an essential part of each unit. The assignment is done in the students' time outside class hours.

C) Reporting on telephone assignment, e.g., information hunt using the yellow pages.

D) Discussion and comparison of observations and informal surveys conducted outside class in the speech community, in Phase IV.

Written Follow-up: Open-ended/unstructured.

A) Student workbook. Questions based on the unscripted conversations of the listening component—students listen to the conversations in the language lab and answer the questions.

PHASE IV: Participant Observation

This phase is the key to the syllabus. It provides the essential link with the speech community outside the classroom. Students are essentially assigned to go out and gather data, to make observations to bring back to class.

Listening Component

Making observations of specific aspects of language use in the speech community and getting into the habit of taking notes.

> ***Some of the things that people do with English***
>
> ***Example:* Unit 3**
>
> **Keep a record of all the introductions you observe for a week (including ones where you get introduced to someone). An ideal approach would be to follow a host around at a party.**
>
> 1. **Describe the situation. Formal/informal? Age of people? Place? Describe the people.**
> 2. **Did the person doing the introductions use first names, nicknames, last names, a title?**
> 3. **Did anyone shake hands? Who shook hands with who?**
> 4. **What did the person say when he/she did the introduction?**
> 5. **What did the other people say when they were introduced?**
> 6. **What happened next? (Did the first person explain who the other two people were? Did everybody talk about something else?)**

Spoken Component

Conducting surveys of a) language use and b) language attitudes. Students ask questions to native speakers in the immediate community.

Written Component

This involves systematic documentation of observations for comparison with other students in class. Questionnaires are provided in the student workbook.

Student and teacher roles

What are the implications of this kind of syllabus design for language methodology? Obviously, the focus on speech events and on working from texts of spoken discourse implies that a certain amount of listening practice will be integrated with spoken practice: a well-established technique in language teaching, but in a different guise, since the main burden of interpretation is on the learner. Initially, the role of the teachers would be to ensure that the students enter into the spirit of the class right from the start: to make it quite clear what kind of approach they will be expected to take toward the language they are learning. Students often come to the language class with a conservative set of expectations about their role and about the kind of learning activities they will be engaged in.

Since the first few classes will set the tone for the course, they should be devoted to training students to listen for variation in language use and to view language as acts of communication rather than as an abstract grammatical system separate from a social context. The following guidelines indicate what kind of direction the discussion in the first few classes should take: (the use of metalinguistic terminology should be kept to a minimum, although one or two key terms/phrases such as *appropriateness*, *communicative function*, *language in use* vs. *language structure* and *speaker* vs. *hearer* could be introduced).

1. What are the functions of language? What kinds of things do we need to communicate to other people?

2. What kinds of actions are only accomplished with words? What kinds of things do we need to do with language? What kinds of things have you already done with language in the situations you've been in today?

3. What kinds of expressions are used to do these things with the language you are learning? Would this always be expressed in the same way? Why not? What kinds of social factors cause people to change the way they make requests, the way they greet each other, the way they express gratitude and so forth?

Throughout the remainder of the course the role of the teacher will then be: 1) to coordinate the language activities inside and outside the class; and 2) to act as a native speaker informant. The role of the student will be to listen, to practice using the language and, above all, to ask questions about the why and the how of doing things with words in English.

Obviously, no functional syllabus can deal with all aspects of language use, any more than a grammatical syllabus can cover all aspects of the structure of a particular language. Course materials can do no more than stimulate and guide the learners in the early stages of a long and complex learning process. At best, teaching materials can provide students with a set of strategies for interpreting the speech events and conversational interactions they participate in outside the classroom setting. Students who are living and learning in the target speech communities can and should be involved in fairly systematic observation of language use in those communities and should be provided with guidelines as to what sort of questions they need to ask.

That is what these materials aim to do. Since so little is known as yet about spoken discourse and about the ground rules for language use in different speech communities, the only option open to the language teacher and syllabus designer concerned with language use is to engage the student in something like an informal ethnography of speaking in the classroom and outside; to substitute interpretation for explanation, since the latter presupposes that we have some established body of knowledge and some set of empirical observations to guide us; and, finally, to center the interpretive activities around a series of texts of spoken discourse which are as authentic as possible.

ESL Speaking Skills in the Elementary Bilingual Classroom

Anna Uhl Chamot

The University of Texas at Austin

Although there is general agreement on the importance of developing speaking skills for the elementary school ESL child, the actual techniques for developing these oral skills to a level of proficiency are frequently left to intuition. Teachers discover that textbook drills do not produce fluent speakers, and that the ESL learner who does become communicatively competent does so through participation in social situations.

Current secondary ESL approaches and elementary language arts materials for native speakers are synthesized in an ESL Speaking Model which associates cognitive processes with different levels of speech behavior.

Suggested criteria are then presented for the selection and preparation of speaking activities. These include: specification of learning objective and student behavior; topic selection that fits the age, needs and interests of children; selection of language structures and vocabulary of greatest utility; timing and pacing of speaking activity; audio-visual aids as speech stimuli; incorporation of other language skills in the activity; comprehensibility-based evaluation of children's oral production.

The major part of this presentation is a demonstration of sample speaking activities for each level of the ESL Speaking Model, with emphasis on approach and techniques for facilitating those higher levels of communicative interaction frequently absent in present commercial elementary ESL materials. The structuring of real social encounters within and without the school setting is described, as well as the use of community resources to foster the development of communicative competence in children learning English as a second language.

Speaking is a critical social skill (Rubin 1975:120; Logan, Logan, Paterson 1972:104). In order to participate as a full member of a social group, an individual's speech needs to fit that group's criteria of both comprehensibility and appropriateness.

Comprehensibility is, of course, essential if communication is to take place. The listener's attention will not last long if the speaker's ideas are not expressed in an easily understandable manner. Very young children learning their native language have comprehensibility problems; often only the immediate family can understand them (Brown 1973:245). By the age of about three, however, they are reasonably successful in making themselves understood outside the family circle (Chastain 1976:331). But even school-age children are still making mistakes in pronunciation, grammar, and word choice—which their peers mercilessly correct. When the school-age child learning English as a second language makes inevitable mistakes in comprehensibility, native English-speaking peers often express complete intolerance. The reaction of the ESL learner may vary from total rejection of the language the peer group represents to a strategy of speaking as little and as infrequently as possible (Rubin 1975:112) or to a determination to master English and be accepted by the peer group. Which attitude is taken is certainly due in some part to the sensitivity of the ESL teacher to the comprehensibility needs of the ESL learner.

In addition to being comprehensible, speech needs to be appropriate to the social situation. A speaker who selects words and expressions that are either too formal or not formal enough for a given occasion will receive a perhaps subtle but nonetheless negative reaction from the listener. Young children learning their native language also have to learn about social appropriateness (Hymes 1974:75). For instance, in languages like Spanish and French, they have to learn by school age that some people must now be addressed as *usted* or *vous* rather than *tu* or *tu*. Children soon learn that some words and expressions which can be freely used with peers brings swift retribution if used in the hearing of teachers or parents; as they grow more sophisticated socially, they learn which speech styles are most effective in different situations.

In selecting appropriateness of language, the ESL learner encounters even greater difficulties than the native speaker. Once a reasonable degree of comprehensibility and acceptance within the peer group has been achieved, the ESL child may not realize that additional styles of English need to be acquired in order to communicate successfully in other kinds of social encounters. The ESL teacher needs to provide experiences designed to help the learner choose the speech style that is appropriate to a variety of situations (Taylor and Wolfson 1978:35).

Although ESL textbooks and curriculum guides provide many exercises designed to develop speaking skills, teachers often find that these do not necessarily lead to real speaking fluency, particularly if the emphasis is on pattern drills (for a neurofunctional explanation of the failure of pattern practice drills, see Lamendella 1979:5-18). At least two reasons can be found for this discrepancy between goals and actual outcomes. First, many ESL speaking activities are in the form of drills which are too often irrelevant to the child's real experience, so the child does not perceive them as in any way related to real communication (Chastain 1976:333). Second, most ESL materials contain extremely few activities designed to bridge the gap between manipulating the language with a fair degree of comprehensibility, and communicating through the language with social appropriateness (Rivers and Temperley 1978:15).

The purpose of the speaking activities described at the conclusion of this paper is to show how the relevance and interest of speaking drills or language-practice activities can be enhanced, and to suggest language-use activities to help the child become communicatively competent through use of appropriate language in different social situations.

The skill of speaking has several different components. In order to communicate an idea, the effective speaker has to select the specific words that convey his or her semantic intention, arrange them and modify them according to rules of grammar, apply social and situational rules to the grammatical sentences, and apply phonological rules that will make the sentence comprehensible to the listener.

The second language teacher helps students acquire these different components of the skill of speaking in a variety of ways. As primary source of linguistic input, the teacher provides a model of speech against which children can match their own production. At the same time, the teacher needs to be aware of other models of English that children hear: in the community, on radio and television, from the peer group. Peer models especially provide input on the kind of language needed for successful social interaction; it may be in the form of a dialect different from the teacher's own, but should not be viewed critically on that account (Chamot 1979). The ESL learner will eventually have to learn a number of styles and registers; usually he/she starts with an informal one appropriate principally with other children of the same age.

1. An ESL Speaking Model

Foreign language methodologists generally propose a hierarchy of speaking skills ranging from easiest to most difficult (Chastain 1976:351-355). Typically, they describe the lowest level of speaking skill as the ability to imitate words and sentences from a spoken model. At a higher level is the ability to manipulate portions of a language segment so that new sentences are formed. The teacher's direct control is lessened as students increase their ability to create their own sentences, so that eventually they can operate successfully in a free communication situation. This last is usually the highest level which the foreign language learner is expected to reach (Chastain 1976:335).

However, for the child learning English as a second rather than as a foreign language, the goal of comprehensible speech used communicatively can be characterized as a midway goal. The native speaker of English comes to school communicatively competent in a particular social environment, and in the elementary school language arts program this basic competence is extended, amplified, and elaborated. The learner of English as a second language must do the same in order to participate fully in both the English speaking environment of school and the larger community.

A synthesis of foreign language speaking skills hierarchies and of the types of speaking activities common to language arts programs can provide a sequence of speaking skills for the ESL child. A proposed hierarchy of speaking skills that provides a sequential plan for teaching speaking in an elementary school ESL program is shown in Table 1.

The first level skill involves imitation of single words, phrases, and sentences. The learner practices pronunciation, intonation, and grammatical structure.

TABLE 1

ESL Speaking Model

Level	Cognitive Process	Student Behavior
6	Discourse	Use appropriate speech style in different social situations. Use appropriate speech style for different purposes.
5	Communication Experience	Communicate in the community. Plan a class activity. Explain a process. Conduct an interview.
4	Guided Communication	Participate in structured discussions. Give directions. Present oral reports.
3	Interaction	Ask and answer meaningful questions. Role play with known sequences. Present recombined information.
2	Variation	Produce recombination sentences. Produce recombination sequences.
1	Imitation	Produce contrasts in words. Produce contrasts in sentences.

At the second level, the learner varies given sentences and longer sequences in order to express different meanings.

The third level requires the learner to use the speaking skill acquired to interact with classmates in several ways: by asking and answering questions that have personal meaning, by using known sequences to engage in role playing, and to present recombined information.

At the fourth level, the learner engages in activities which have been structured by the teacher in order to elicit communication in the second language. Various communication games and activities are a feature of this level.

By the fifth level, the student is expected to use the communicative skills practiced in class in the larger social context of the community (as proposed by Corder 1977:10), and to use speaking skills in class without elaborate structuring by the teacher.

The sixth level calls for the student to make appropriate shifts in style and register depending upon the variables of social situation and purpose of speech. At this stage, the teacher's task is to sensitize students to the importance of appropriateness in speech style (Rivers and Temperley 1978:11), but real acquisition of a variety of speech styles must necessarily occur individually as the student encounters different social situations beyond the classroom.

The student behaviors described on the Speaking Model represent some of the types of activities that are typical of the cognitive processes specified, but are not intended to be all-inclusive. This Speaking Model can be used to select and classify different kinds of speaking activities so that a teaching/learning sequence is established. The levels are generally hierarchical, that is, each builds on the preceding ones. However, there will usually be some elements of the behaviors described at the Discourse Level present in lower levels because examples of socially appropriate speech will inevitably form part of the child's linguistic input from the very beginning. The hierarchical arrangement does not imply that all speaking activities in a class will be at a single time; on the contrary, only newly introduced language segments will be practiced at the first two levels, for as soon as children gain familiarity with a segment, they should engage primarily in activities at levels three and above.

2. Criteria for Speaking Activities

Criteria for selection of speaking activities need to be established so that maximum learning takes place. The first important criterion is the identification of the learning objective (Logan, Logan, Paterson 1972:107). By specifying the expected outcome of a given speaking activity, the teacher guards against those which are merely busy work or recreational. The Speaking Model can be used to clarify learning objectives by identifying what the child will do during an activity, and then matching the behavior to the cognitive process it reflects. For instance, if children will be learning a dialogue, then the learning objective will be found at Level 1 of the Speaking Model, but if they will be improvising a dialogue using known words and structures, then the learning objective will be found at Level 3. Objectives at the higher levels of the Speaking Model emphasize communication of message rather than correctness of form. Learning objectives need to be communicated to the children as well. When they know which speaking skills are being practiced, they will be more likely to monitor their own output effectively.

A second important criterion has to do with the content of the speaking activity. Content should be relevant to the age of the child; although an older beginner may need the same kinds of practice that the younger one does, the material that is used for practice should be more sophisticated.

A third criterion involves the length of the speaking activity. In general, speaking activities on the lower levels of the model need to be kept quite short and repeated after intervening activities of different types. Practicing at the manipulative level of language easily becomes boring, so the teacher needs to vary the type of activity frequently. Communicative activities at the higher levels of the model can last considerably longer because the focus is on content rather than form, and children's interest is thus more easily maintained.

A fourth criterion has to do with the model of linguistic input on which the child will be basing his or her own output. Speech input should be clear, easily understood, and free from distractions such as mannerisms, overuse of idiomatic expressions, incorrectness, and unnatural pace. Teachers need to monitor their own speech to be sure that it is an acceptable model for children learning English (Rubin 1975:109).

The fifth and most important criterion has to do with the meaningfulness of the speaking activity. Although meaningfulness is built into the higher level communicative activities, this is not always the case with the language-practice activities at the lower levels of the Speaking Model. Too often manipulative exercises become mechanical and detached from even the most tenuous contact with realistic communication. In order to prevent this from happening, a context for every speaking activity needs to be provided. This means that every sentence for repetition and every variation drill needs to be set either in a real situation or an imaginary one where the utterances being practiced could occur naturally. A real situation can often be established through a game which establishes a purpose for language-practice drills. An imaginary situation can be set up by the teacher with context provided through description, pictures, and real objections. When manipulative level activities are thus contextualized, children can more easily perceive the language being practiced as a real part of the communicative process.

The sixth criterion is an outcome of the first, for the learning objective set for a speaking activity will determine how it is to be evaluated. If the learning objective is at Level 1, then the teacher will evaluate pronunciation and intonation. At the next two levels the teacher may be concentrating on grammaticality and vocabulary, while at the communicative levels, the focus will be on content and meaning rather than on form. The same type of speaking activity may be evaluated for different aspects on different occasions. While comprehensibility may be the main objective for a considerable time in the child's acquisition of English, the further goal of social appropriateness also needs to be evaluated.

Criteria such as these help teachers plan speaking activities that can become a purposeful and central part of the ESL curriculum.

3. Sample Speaking Activities

Representative speaking activities for each level of the ESL Speaking Model are described in this section. Except for two general suggestions for activities at Level 1, all of the specific activities described call for individual, paired or small group speaking experiences. These call for classroom organization in which the teacher plans for several different things to be going on simultaneously, rather than having total group practice of identical spoken patterns. Although this type of teaching requires detailed planning, it allows for more individual responses and more frequent opportunities for all the children to engage in group conversational practice (Chastain 1976:357). The teacher's role becomes that of model, initiator and facilitator, and student-student interaction is actively cultivated. When the teacher organizes language learning experiences systematically, an environment is established in which children can engage in the creative construction process of second language acquisition (Garcia and Gonzalez-Mena 1976:4).

LEVEL 1—IMITATION

When imitating the teacher's model, the child needs not only to understand the meaning of the language segment but also to feel that there is a good reason for imitating it. This can be difficult to provide, for in real life we do very little oral repetition of what someone else says.

Suggestions for making group repetition more enjoyable can be found in innovative approaches such as chanting (see Graham 1979) and in choral speaking activities using nursery rhymes and children's poetry (see Rubin 1975:123-128). Material for choral speaking for the ESL child, however, needs to be selected carefully so that it is not only meaningful but also has standard spoken syntax and useful vocabulary (i.e., poetic license, archaisms, and rare structures are not useful models for beginning ESL speaking practice).

For individual repetition, small group work is essential. The teacher working with six or seven children can provide a meaningful context, encourage each child to participate, and give individual attention to each child's spoken output. The most fluent children will be asked to imitate the teacher's model first, thus serving as additional models for the less fluent children.

One suggestion for younger children is the use of a toy animal or puppet. The teacher starts the imitation activity by showing the toy, and the speaking sequence goes something like this:

Teacher: This is my pet dog, Snoopy. He doesn't obey me! When I say, "Snoopy, sit down," he doesn't sit down. Do you think Snoopy will obey you? Let's try! I will tell you what to say. Ready?

Child: (Nods.)

Teacher: Snoopy, stand up. (Keeps toy dog in sitting position.)

Child: Snoopy, stand up. (Teacher makes dog stand up.)

Teacher: Snoopy, lie down and turn over. (Keeps dog in standing position.)

Child: Snoopy, lie down and turn over. (Teacher makes dog do so.)

After the initial practice, a child can manipulate the toy, having it move only when another child imitates the teacher's model.

A suggestion for older children requires the preparation of a short tape with either spelling words or a narrative on it, and with repetition pauses after each segment. A group of children sit at a table with pencil and paper or worksheet. One child listens to the tape through a headphone, and repeats each word or phrase on the tape during the pause provided. The other children must write down what is being dictated to them. In the case of a longer narrative, this may be a cloze dictation in which children have a worksheet with part of the text on it and must fill in the missing words as they hear them. Different children can take turns dictating the taped material. This activity motivates children to repeat as accurately as possible in order to help their peers write the exercise correctly.

LEVEL 2—VARIATION

At this level children are encouraged to vary sentences by adding or changing parts of them. Here again, contextualization and individual responses are preferred so that children can make meaningful rather than mechanical variations.

Younger children can derive continuing enjoyment from a mystery box with ever-changing contents (Rubin 1975:141). The teacher can review vocabulary as different objects (realia, toys, miniatures, doll house furniture, etc.) are dropped into the mystery box. Then children take turns dipping into the box and having others guess what each is holding inside the box. The dialogue can go something like this:

Child 1: (Dips into the mystery box and picks up an object without showing it.) Guess what I have!

Child 2: Do you have the car?

Child 1: No, I don't.

Child 2: Do you have the chair?

Child 1: Yes, I do.

For older children, a strip cartoon can be modified to serve as a context for variation practice. Choose a cartoon sequence that has a question-and-answer dialogue, and cut away either the questions or the answers. It may be necessary to re-write the remaining text so that the structures and vocabulary are known by the children. A transparency of the modified strip cartoon is then shown, and children are asked to provide the missing bits of dialogue from the clues given by the drawings and the text.

LEVEL 3—INTERACTION

At this level, the emphasis is on interaction between children, and groups or pairs can work together at the same time (see Mellgren and Walker 1979: P-6).

Children can learn to ask and answer meaningful questions that are personalized. The teacher can suggest, orally or in writing, three or four questions, such as:

When is your birthday?

What do you want for your birthday?

What are you going to do on your birthday?

Each child then interviews another child in the classroom, and when all the interviews have been completed, children take turns presenting the information discovered to a group.

Simple riddles can also be used for interaction practice. Humor is a difficult aspect of a second language to acquire, but simple plays on words and jokes (the cornier the better) can be a good beginning point (Trachtenberg 1979:91). Many riddles follow a very simple format, and once this is learned, children with even limited knowledge of English can invent their own and have their classmates guess the answers. The teacher may need to rewrite riddles to make the structure and vocabulary within the reach of the ESL child, but many of the classic descriptive riddles ("What is [*adj.*], and [*verb*]?") can be used without simplification.

LEVEL 4—GUIDED COMMUNICATION

At this level, children engage in discussions and problem solving activities that are structured by the teacher. In communication activities, the content of the message being transmitted is more important than its form, hence teacher corrections should not be given during the activity (Chastain 1976:337). Many of the discussion activities recommended for regular language arts classes can be used at this level.

Using a map to give directions from one point to another is an activity that can fit into the needs of both younger and older children. The map can be of the classroom, of the school, of the neighborhood around the school, or of the whole town. It can be reproduced on the chalkboard, a transparency, a poster, or a playing board. It can even be drawn in chalk on the floor of the classroom so that children can walk from one point to another as they give directions. A neighborhood map can be used to give directions to and from the home of each child, and a city map can give children practice in learning how to get to major points of interest.

An activity for second grade and above has the teacher ask children to think of a place each would like to visit, but to keep it secret. Each child then writes down five things needed for the trip. The class is divided into small groups and children take turns telling what they will need. Others in the group ask why each thing is needed, and an appropriate reason has to be given. After all five things have been questioned and explained, group members begin guessing the destination (Yorkey, Barrutia, Chamot et al. 1978:429).

Group decision making activities encourage participation in discussion (Mountain 1974:24) and in the ESL class can provide speaking practice in a natural context. In this type of activity, a problem is first set up by the teacher. Individual children record solutions to the problem, then the class is divided into small groups which discuss the individual solutions and eventually reach a group consensus. Evaluation of the solutions is provided through an authoritative source, and children usually find that the group decisions are closer to the correct answer than are most individual solutions. As an example, the different steps of a recipe for brownies can be presented in incorrect order, and the task is for children to arrange them in correct order. After individual children have put the steps in order, they join a group to discuss a final order. The teacher can then present the actual recipe for comparison, and as a follow-up the class can engage in an actual cooking experience.

A guided communication activity for middle grade students with some reading ability in English is called Information Exchange. In this activity, the class is divided in half. One half works on an assignment with the teacher or individually. The other half is given a worksheet to read carefully; the content of the reading passage can deal with any subject area relevant to the class (science and social studies passages are good choices). After the material has been read, the teacher collects the worksheets and each student who has read the information must now report it to one who has not. After the paired discussions, students who did not read the information are given a worksheet with questions to answer about the material, while the teacher works with the other half of the class. This activity encourages students to report accurately and to question anything not understood clearly.

LEVEL 5—COMMUNICATION EXPERIENCE

Activities at this level provide experiences in which children use English for communication in a real, rather than a contrived, situation. The ideal experience is for children to engage in communicative encounters in English in the community. The teacher can encourage this type of encounter by keeping a Communication Journal for younger children, and by having older children keep their own journal. In it will be recorded the date and description of the communicative encounter, with comments on its degree of success. These journals should be shared with the teacher and with other children on a regular basis. A Communication Journal can help motivate a child to actively seek to use English outside of school, and can provide the teacher with insights into the child's ability to cope in a free communication situation.

Classroom centered communication experiences can center around group planning for a class activity such as a party or picnic. Children can volunteer to work in groups to plan each aspect of the class activity before the final plan is synthesized and implemented.

Another communication experience which can take place in or out of the classroom involves conducting an interview with a person whose job or hobby is of particular interest to the group. Suggestions for interviewing preparation, techniques, and follow-up can be found in most language arts teaching guides (see Logan, Logan, and Paterson 1972:118). The basic procedure calls for selection of the person to be interviewed, writing the interview questions, and practicing interview procedure.

Many communication experiences draw on the individual interests and talents of children. An example is one in which each child explains a process to a group of classmates. Discussion will yield information about special abilities, which can range from building models, cooking, playing an instrument, ironing a shirt, making a toy, etc. Each child should bring any materials necessary to explain and demonstrate how to do something that he or she does well. In addition to providing speaking practice in English, this activity fosters a positive self-concept and pride in achievements that are not necessarily school-related.

LEVEL 6—DISCOURSE

As mentioned earlier, there are elements of discourse at every level of the

Speaking Model, for as soon as a speaking activity is contextualized, the language used in it will be appropriate to the social situation in which it occurs.

Role-playing experiences can help children become aware of different speech styles and registers to use according to the situation portrayed. In order to set up realistic situations for role-playing, the teacher needs to be informed about the kinds of social encounters the ESL child will typically have to deal with in a given community. Useful suggestions for role-playing situations can be found in language arts methods texts (see Rubin 1975:132-134), but to these the teacher should add situations that are part of the daily life of the ESL child. Familiar situations might include: being lost and having to tell an adult; damaging or breaking something and having to explain and apologize; being sick and having to go to the doctor, or accompanying another family member to the doctor and having to explain the nature of the illness; getting in trouble at school and being sent to the principal; asking a policeman for help; using language to avoid difficulties with policemen and other authorities; watching television at a friend's house; getting into a fight with another child who speaks only English; babysitting; buying something; getting a prescription filled; telephoning for information. After deciding on the situation, children choose the characters that will take part in it. The teacher should discuss with children the feelings of the characters involved in the role-playing situation, and their probable style of speech, whether formal or informal, friendly or antagonistic, simple or elaborate. Props add greater reality to role-playing and also help children enter into the characters of the persons they are portraying. These can be as simple as a hat, badge, handbag, briefcase, stethoscope, whistle, notepad, or toy telephone. After the situation is role-played, children can discuss the effectiveness of the kind of speech style each character used and make suggestions for changes.

As children become aware of the uses of different styles of speech, they will begin to use language not only to communicate ideas and feelings but also as the means to social survival.

Teaching ESL Technical Writing With the Personalized System of Instruction

Gayle L. Nelson and John B. Keenan

West Virginia University

Increasing numbers of international students are coming to the United States to study Engineering; in fact, Engineering is the most popular field of study among international students at American colleges and Universities (IIE Reports, 1978). This increase creates a need for effective materials and methods for teaching technical English to international students.

This paper describes an application of the Personalized System of Instruction (PSI) to the teaching of technical writing to international students. PSI, as used in this course, is distinguished by: 1) mastery learning and the use of small units which specify what the student is to learn, 2) unit study guides which include writing assignments, 3) frequent evaluation, 4) self-pacing which allows individualized time for learning, and 5) the use of proctors to maximize individual attention and assistance available to each student.

Engineering is the most popular field of study among international students at American colleges and universities (IIE Reports 1978). Between 27% and 34% of all engineering Ph.D.'s from American academic institutions are earned by international students (*The Chronicle of Higher Education*, 1978). West Virginia University is one of the many universities that reflect this increase in international students, with approximately 350 international students of engineering, about half graduate students.

In the summer of 1977, the University investigated the feasibility of expanding its EFL program to accommodate the growing number of international students. Faculty and international students were interviewed regarding ESL needs. In-depth interviews were conducted with the engineering faculty and students due to the large number of international students studying the subject. The faculty consistently commented on the international students' low level of English skills, particularly their low level of writing skills in lab reports, research papers, letters, and theses. The international students also admitted having writing difficulties, especially those of conciseness on essay exams and lab reports.

In the fall of 1977, representatives from each engineering department met with the ESL faculty to confer in the development of a technical writing course. The engineering faculty supplied copies of student reports, course requirements, and textbooks; all were used in the course. They also reviewed technical writing textbooks, commented on course construction, and eventually advised their students to take the course. In selecting a course design, two major concerns were: 1) meeting the individual needs of students with differing English levels and problems; and 2) directly relating the technical writing course to students' engineering courses so that it would provide a quick and noticeable improvement in the quality of their writing. It was primarily because of these two concerns that the Personalized System of Instruction (PSI) was selected.[1]

At that time, many of the additional benefits of using PSI in an ESL class had not been fully realized.

PSI, as used in technical writing, is distinguished by: 1) mastery learning and the use of small units which specify what the student is to learn; 2) unit study guides which include writing assignments; 3) frequent evaluation, 4) self-pacing which allows individualized time for learning; and 5) the use of proctors to give individual attention and assistance to students' particular problems, the first concern of differing English levels. The self-pacing component of PSI also allows for differing English levels; the advanced student can finish early and the less-advanced student can take more time and receive additional help from the proctor. In writing the study guides, material from the WVU College of Engineering was used so that the questions and writing assignments related to the students' engineering courses, thus meeting the second concern. Finally, the unit mastery component contributes to both concerns by insuring that each student attains a high level of competence with the subject matter in each unit and eliminating the problem of cumulative failure, encountered when students' failure at early course material inhibits their performance on later material. Given the concern for rapid, noticeable improvement, the requirement of unit mastery provides an assurance of improvement in quality through the high mastery criterion and frequent performance evaluations.

PSI has been the topic of hundreds of published research articles and the overwhelming majority of these studies provide convincing evidence that PSI has been significantly more effective than traditional approaches. Several reviews have indicated that average student performance ranges 9-15% better than student performance in traditionally taught courses (Hursh 1976, Kulik, Kulik, and Smith 1976). Research has demonstrated that students in a PSI course retain more material when tested months later. In addition, it has consistently been shown that although students report that a PSI course requires more work, they also report enjoying the course more than others (Kulik, Kulik, and Smith 1976). PSI has been used successfully in a wide variety of disciplines, from the "hard science" disciplines of engineering and physics to the abstract fields of poetry, philosophy, and English (Hursh 1976).

1. Course Construction

The technical writing course is divided into twelve units—four grammar units and eight units based on the text, *Reporting Technical Information* by Kenneth Houp and Thomas Pearsall (1977). Each unit is composed of a reading assignment, 20-30 study questions on the reading, appropriate writing assignments, and three equivalent forms of a unit quiz. After students complete both the written assignments and the study questions, they take a quiz. A score of 90-100% on the quiz, indicates mastery of the material reflected on a cumulative progress chart which is attached to the students' quiz folders. If they do not obtain 90%, they resume studying and when ready, take a different form of the quiz without penalty (Figure 1).

---

[1] For additional information, contact the Center for Personalized Instruction, Georgetown University, Washington, D.C., 20057.

2. Study Guides

In using PSI to teach ESL, quiz mastery marks a student's passage to the next unit, but it is not the sole focus of the course. An essential component is the student's written work in the study guide. Each study guide contains 20-30 questions and problems from the unit reading assignment that necessitate student implementation of material, not memorization. An example of such implementation is:

> Your company has recently developed a new deep water oil rig which utilizes a completely new system of drilling. Your job is to write a report on the subject which will be read by the president and board of directors. How much can you assume these people know about deep water drilling? Why? What specific things about the new oil rig would you want them to know?

Each study guide also includes several writing assignments that students complete and discuss with a proctor before taking the unit quiz. The following writing assignment from the unit on prose elements of a technical report serves as an illustration:

> 1. On a separate sheet of paper, write a letter of transmittal to one of your professors to whom you are submitting a paper. You may use the form in the textbook or check the form that your department recommends. Show this letter to a proctor before taking a quiz. The letter should be neatly written.
> 2. Write a descriptive abstract for a paper that you have written. You may refer to your departmental guidelines. Show the abstract to a proctor before taking a quiz.
> 3. Write a statement of purpose for a report that you are writing or for a report that you plan to write. Show the statement of purpose to a proctor.

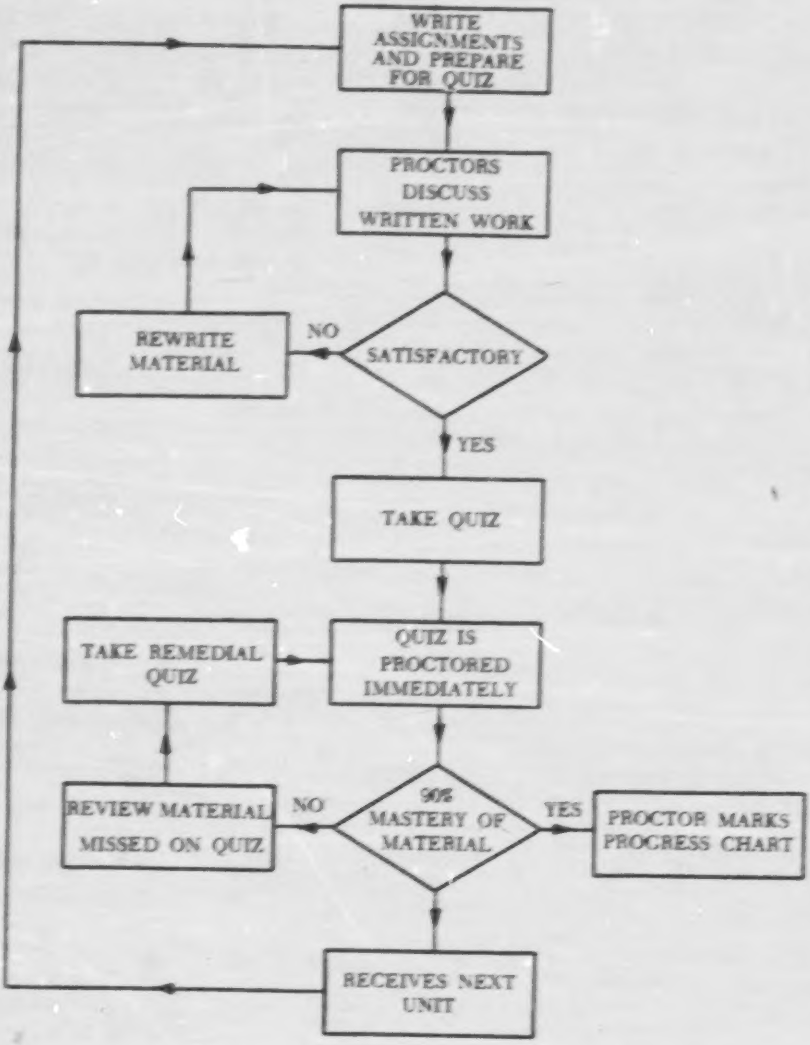
The study guides move from tight control to little control of student responses, from rule identification to rule implementation, from defining an abstract to writing an abstract.

3. Proctors

The proctors are graduate and undergraduate native English speaking students who are studying to become ESL instructors. They are reimbursed with credits in fields such as ESL Methodology, English, or Education. As shown in Figure 1, the proctors play a vital role in a PSI writing course, with three main responsibilities: 1) effectively discussing student work, 2) correcting quizzes, and 3) recording progress on the students' progress chart.

Discussing their written work with the students is the most important and

Figure 1. Procedure for Student Progress Through Course.


WRITE ASSIGNMENTS AND PREPARE FOR QUIZ
PROCTORS DISCUSS WRITTEN WORK
NO
REWRITE MATERIAL
SATISFACTORY
YES
TAKE QUIZ
TAKE REMEDIAL QUIZ
QUIZ IS PROCTORED IMMEDIATELY
NO
REVIEW MATERIAL MISSED ON QUIZ
90% MASTERY OF MATERIAL
YES
PROCTOR MARKS PROGRESS CHART
RECEIVES NEXT UNIT

flow chart

most difficult aspect of the proctor's task. Rather than simply providing the correct answers, their role is to ask questions that help the students arrive at the corrections themselves. Examples of such questions or responses are:

> Can you think of another way to say this? Read this carefully and try to figure out what's wrong. What tense are you using? Is this word countable or uncountable? Check the spelling of this.

Rather than simply telling the student the correct answer, the proctor prompts the student to supply the answer, thus considerably improving student performance and retention of material (Johnson 1978).

Proctors also make note of ambiguous study guide and quiz questions, rewriting them for the next class; the course is continually being revised. They also refer students to the text or to supplemental material when necessary. An additional function of the proctor is to help students with technical reports they are concurrently doing in their engineering classes. If proctoring is to be effective, the proctor/student ratio needs to be small; about 5:1.

4. Conclusion

The success of the technical writing course can be measured by student and faculty evaluations, pre-test and post-test scores, and improved writing in engineering courses. The course has been taught three times so far, with approximately 20 students in each class; student evaluations have been consistently high. The students are very enthusiastic about the method, the self-pacing, the large amount of information they learn, and especially the learning of various forms of writing—types of letters and reports. These high evaluations are consistent with research on PSI: Kulik, Kulik, and Smith (1976) report that students rate PSI courses significantly higher than traditionally taught courses.

The pre-test and post-test includes one question from each of the twelve units and asks the students to write a letter applying for a graduate assistantship. A simple t-test for repeated measures indicates a significant increase in scores between the pre-test and the post-test ($p < .01$). In comparing the first and second letters, the improvement in overall writing skills is immediately noticeable: there are fewer errors in grammar, fewer mistakes in the use of tense, agreement, articles, prepositions, even spelling. Also noticeable is the use of conventions, such as the appropriate form of the letter, abbreviations, capitalization, dates, spacing, and style.

At the completion of the first course, engineering faculty were asked to comment on any improvement in student writing; the replies were overwhelmingly positive. The Department of Chemical Engineering, for instance, began requiring all their international graduate students to take the course and other engineering departments are considering the same.

An additional feature of this endeavor warrants further comment. The success of the PSI technical writing course at West Virginia University is in large part a result of the collaboration of three typically unrelated fields: English as a Second Language, Engineering, and Instructional Psychology. By incorporating information from each of these three disciplines, the content, design, and relevance of the course was greatly improved. This inter-disciplinary approach seems particularly important to university ESL instructors who are teaching English to students whose primary field of study is something other than English.

Rhetorical Competence and EST Discourse

George R. Hepworth

Bilingual Education for Choctaws of Mississippi

ESL teachers are aware that effective communication often depends on recognizing when the actual purpose of an utterance is different from that suggested by its surface form. Implicit functions are, therefore, a frequent source of difficulty for ESL speakers. Unfortunately, however, the importance of this fact is too often overlooked for written discourse. Texts used in teaching English of Science and Technology (EST), for example, tend to focus on the specialized vocabulary rather than the rhetorical structures and functions of EST discourse.

This paper presents the results of an analysis of EST discourse using a rigorous method of description—the Discourse Bloc Model, a model which identifies units of discourse by their functions rather than by their forms. When the Introductions to Experiment Reports from several fields of science were looked at, the resulting description identified the rhetorical functions employed consistently. An idealized version of the Introduction to an Experiment Report can thereby be proposed. The predicted correspondence between implicit, rhetorical functions and explicit discourse functions also was found to exist. The term *rhetorical competence* is used to refer to the ability to recognize this use of discourse functions in EST.

Communicative competence has become one of the more important ideas in ESL pedagogy in recent years (Paulston 1974, Rivers 1976). In one excellent definition of communicative competence, Savignon (1972) says:

> Communicative competence may be defined as the ability to function in a truly communicative setting, that is, in a dynamic exchange in which linguistic competence must adapt itself to the total information output, both linguistic and *paralinguistic*, of one or more interlocutors. (Emphasis my own.)

This definition, which focuses on the essential role of paralinguistic factors in the proper interpretation of some utterances, also serves to illustrate the current preoccupation with spoken, as opposed to written, discourse in most discussions of communicative competence.

In this paper, I discuss written discourse, more particularly discourse written in the English of Science and Technology (EST). I first attempt to show that, in much the same way that oral communicative competence often involves recognizing the difference between the explicit functions of certain utterances and the implicit functions understood for those utterances, the rhetorically competent reader of EST discourse recognizes when certain explicitly-signalled discourse functions, such as coordination, have other, implicit

---

I wish to thank John Lackstrom and Will Pitkin for valuable comments on an earlier version of this paper. Any errors or deficiencies are, of course, my own.

functions necessary for the proper interpretation of that discourse. I will introduce a method of discourse analysis known as *Discourse Blocs*, originally developed by Will Pitkin (1969, 1973), in order to provide a framework for this discussion.

1. Implicit and Explicit Functions

One factor in the proper interpretation of utterances is the recognition of implicit functions of certain sentence types. Sentences of the type "Do you have X?" have in normal conversation the force of *requests*. Consider (1) and (2):

(1) Do you have the correct time?
(2) Hey, got the time?

For our purposes, it is not necessary to indicate the exact ways in which (1) and (2) are understood as requests and not as yes/no questions, that is, to have implicit communicative functions.[1] However, it is necessary to realize that the communicatively-competent speaker of English would reply to (1) and (2) with (3) and not with (4):

(3) Yes, it's about 9:30.
(4) Yes, I do.

2. The Functions of Written Discourse

Our hypothesis is that in some written discourse, such as that of EST, there is an analogous situation. Explicit discourse functions—such as causality and contrast—may also serve other, implicitly determined, rhetorical functions specific to the type of discourse in which they appear. For example, Lackstrom, Selinker and Trimble (1970) listed the rhetorical functions for an experimental report as the Statement of the Problem, the Hypothesis, the Deduction of the Consequences, Description of Procedures, and so on. However, these rhetorical functions are very seldom explicitly identified in the text of any report, so that identifying them as such appears to be a post hoc operation. We wondered if there might not be some way to identify these rhetorical functions by the explicit discourse functions employed, in addition to the content which might indicate those functions. We wondered if the use of causality, contrast, time or space order, and so on, could be correlated with specific rhetorical functions such as the Statement of the Problem, or the Hypothesis (see Figure 1).

In order to test our hypothesis, we carefully analyzed the introductions to several experimental reports. Because we were looking for functional, not formal, properties of this discourse, our analysis had to proceed along functional lines. Discourse Bloc analysis, described below, provides a convenient framework in which to describe both the explicit discourse functions of EST and their implicit rhetorical function counterparts.

---

[1] See Searle (1975) for an excellent presentation of indirect speech acts.

Figure 1.

Oral Communication:	
Communicative Function	= Sentence Type
REQUEST FOR X	"Do you have X?"
Written Communication:	
Rhetorical Function	= Discourse Type
PROBLEM	??
HYPOTHESIS	?? (Contrast, causality, time ?? and space order, etc.)
DEDUCTION OF CONSEQUENCES	

3. Discourse Blocs

Traditional treatments of discourse have been deficient because they are primarily descriptions of the formal properties of sentences, and have often failed to account for the functions of those sentences. Pitkin (1969, 1973) recognized the limitations of conventional analysis and proposed that the defining characteristic of discourse units be a functional one rather than a formal one.

3.1. The Functional Units of Discourse (Defined). Formal units of discourse—the clause, the sentence, the paragraph, and so on—tend to be inadequate descriptions of the basic functions of discourse, even when they are reinterpreted along more functional lines, as in Longacre (1970, 1976). The following discourse shows this.

(5) a. John is a good skier. He doesn't have time to compete professionally, however.

b. John is a good skier; however, he doesn't have time to compete professionally.

c. John is a good skier, but he doesn't have time to compete professionally.

d. Although he is a good skier, john doesn't have time to compete professionally.

e. A good skier, John still doesn't have time to compete professionally.

Although (5a) and (5b) might be considered stylistic variants, (5c) quite clearly could not be equated with them on the basis of its formal properties. However, it is equally clear that the relationship between the two parts of each version of (5) is constant; after a concession is allowed (John's ability to ski well), a contrasting assertion is made (Johns' lack of time). Note that by defining one unit as a concession and the other as a (contrasting) assertion, it is possible to portray the functional relationship between the two units of (5a) in the same way as for (5e). Further, these two units can be said to appear in a contrastive relationship solely by virtue of the fact that they are used together. Concession/assertion requires two units of discourse, and it requires them to appear as parts of a single relationship which can be defined by the functions of those two units. This allows us to ignore, temporarily, the grammatical form of such sentences, while we concentrate on their functions.

There are a number of such functional relationships in discourse. The following is a brief list of those that appear most frequently.[2]

COORDINATE/COORDINATE:	A concatenation of two or more independent units, either additive *X/and Y*, or alternative *X/or Y*.[3]
INCLUDER/INCLUDED:	These are relationships in which a more inclusive unit is paired with one which is more specific. There are four ways in which the shift can be made—*X/namely Y*, *X/for example Y*, *X/including Y*, and *X/especially Y*. Further, there are four modes of analysis—*operation/stages* or *phases*, *genus/species* or *species/specimen*, *complex/components*, and *whole/part*.
CONTRAST/CONTRAST:	There are two main types of contrast—*concession/assertion (true X/yet Y)*, and *negative/positive (not X/rather Y)*.[4]
CAUSE/EFFECT:	Although causality is too complicated to treat adequately here, we can state that, as is the case with other functional relationships, *cause and effect* units appear in discourse. Four broad categories are the *conditional (if X/Y)*, or *unconditional (because X/Y)*, and the *mental*, or logical, cause and the *physical* cause.
ASSERTION/REASSERTION:	A restatement of an assertion appears as *X/that is Y*.
FOCUS/ASSERTION:	A presentential modifier may be used to focus on or to identify the time, place, or source of an assertion.

---

[2] See also Pitkin 1977a, 1977b, or Hepworth 1978.

[3] The X/Y formulas are a shorthand for the functional relationships. X is the first unit in a discourse, Y the second. The type of relationship is indicated by the use of a signal such as "and" or "or".

[4] Since this paper was originally written, I have received by private communication Borkin's (1978) paper on dissonant contrast. It seems to support the general claims being made here. However, it seems best to delay full incorporation for the present.

ASSERTION/INTENSIFICATION ASSERTION/REGRESSION:	An assertion can be made stronger, *X/indeed Y*, or weaker, *X/at least Y*.
COMPARISON/COMPARISON:	Comparisons, or similes and metaphors, are represented as *just as X/so Y*.

Although this list is not complete, it captures the functional basis of the binary discourse units and indicates the broad categories of relationships which appear in all discourse.

3.2. *Hierarchical Ordering.* In Discourse Bloc analysis, all discourse is considered to be hierarchically ordered into increasingly more complex blocs. This ordering appears in (6).

(6) She fixed dinner—making salad, broiling a steak, and cutting and toasting garlic bread—and called her husband.

In this operation *(preparing for dinner)*, the two main stages are *fixing dinner* and *calling her husband*. Fixing dinner requires three sub-stages to complete: *fixing salad*, *broiling steak*, and *fixing garlic bread*. And *fixing garlic bread* in turn requires two sub-sub-stages to complete: *cutting (garlic bread)* and *toasting garlic bread*. The functional hierarchical structure of (6) is reflected in the restraints on formal arrangement of the units. In longer blocs of discourse the same hierarchical ordering of units occurs. As we noted above, these units may be of any grammatical form, so that a single noun phrase, for example, may be used in contrast to several sentences. This property of discourse makes it necessary to redefine its basic units in functional terms. A *discourse bloc* is a functional unit consisting of a main binary relationship between two sub-units; each of those sub-units may be composed of further binary sub-parts. Hence, the absolute size of any discourse bloc is limited to the length of discourse one chooses to consider.

3.3. *Diagramming.* A method of diagramming which reflects the functional and hierarchical structuring of discourse is an essential part of the analysis. These diagrams depict the horizontal and vertical relationships which hold for a given discourse bloc. Figure 2 is the Discourse Bloc Diagram of (6).

The horizontal lines reflect the relationship at each level; the vertical lines topped by a bloc signal such as *and* or *or* indicate which units are involved in the functional relationship at that level and what the relationship is. For example, the two main stages of (7) are *coordinate (X/and Y)*. The first main stage is an operation and three of its sub-stages, so the next level down in the diagram states that there is an operation and stages *(operation/including stages)*. And the third sub-stage thus identified is further broken down into its own two sub-sub-stages *(X/and Y)*.

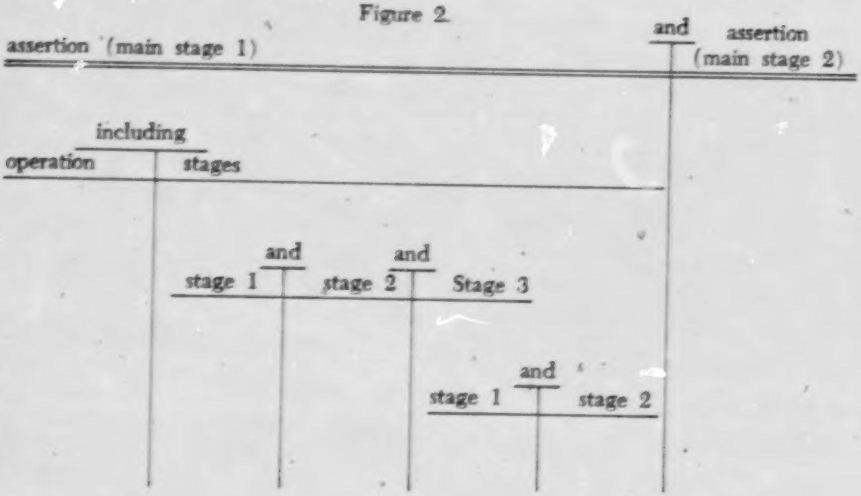
In this rather brief treatment of the Discourse Bloc method of analysis, we have tried to indicate its three main features: 1) the recognition of explicit discourse functions in all discourse; 2) the hierarchical ordering of all discourse; and 3) Bloc Diagram functions.

4. The Analysis of EST

When the introductory paragraph(s) of experimental reports are examined to determine their rhetorical structures, certain similarities appear. As indicated above, Lackstrom, Selinker and Trimble (1970) state that the Introduction to an experimental report will contain: 1) the Statement of the Problem (PROB); 2) the Hypothesized Solution (HYP); and 3) the Deduction of the Consequences (D of C). It is no surprise that these, too, are functionally defined units.

These rhetorical functions do not appear in the inventory of discourse functions found in general use precisely because the EST functions are defined in terms of the specific area in which they are used—namely, the introduction to an experimental report. While the rhetorically-competent discourse analyst and the rhetorically-competent scientist may be able to recognize these features, and perhaps even to use them in writing effectively, it is not at all clear that the same will be true for many ESL/EST students. It is, in fact, quite likely that the ESL/EST student will need considerable practice in learning to use these rhetorical features in writing. It would, therefore, be quite valuable to have a more precise description of the ways in which Problems, etc., are written. Thus, based on the hypothesis suggested above—that there will be a correlation between the explicit discourse functions such as causality and the implicit rhetorical functions of EST such as causality and the implicit rhetorical functions of EST such as PROBLEM—analysis of the introductions from several experimental reports was carried out following the discourse bloc model. As predicted, there are some highly significant similarities, as well as some im-

Figure 2



portant differences. Figures 3 and 4 depict the two basic versions of what we will come to call the Idealized Introduction to an Experimental Report.

Figure 3.

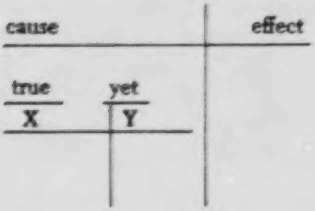
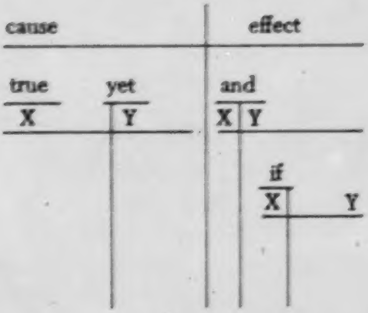


Figure 4.



In these Bloc Diagrams, only those discourse relationships which play a significant and consistent role in the development of the introduction are shown. Also, since Figures 3 and 4 are syntheses of many such diagrams, it is highly unlikely that one will encounter any single introduction exactly like either, although, of course, it will be very similar to one or the other. Note that in most particulars, Figures 3 and 4 are identical, except that Figure 4 has a more elaborated second stage.

After we found the Bloc Diagrams of Figures 3 and 4, we related the explicit discourse functions that appear in Figures 3 and 4 to the rhetorical functions PROB, HYP, and D of C. Problem, being a single rhetorical function, had to be redefined for this purpose. Discourse relationships are binary, which means that the Problem statement must be further analyzed either into two constituent parts or paired with a second related unit (either horizontally or vertically expanded in the Bloc Diagram). Traditional Discourse Bloc analysis did not provide for this description of the problem, although there is some discussion (Pitkin 1977b) of the need to consider it. On the other hand, traditional EST discourse analysis does provide terms that appear to fit into the needed pattern—Problem and Solution. Looking at Figure 3, we intuitively feel that *cause/effect* fits nicely with what we might call *problem/solution*, provided that the nature of problems in science are carefully considered.

Normally, *cause/effect* would mean that either a physical causality or a mental deduction exists, as in (7) and (8).

(7) Because "X" exists, "Y" also exists.
(8) Because "X" exists, we must assume that "Y" exists.

When we are talking of scientific investigation, it is not sufficient to state simply that a solution "Y" exists, nor that we assume it must exist. Rather, it is assumed that several possible solutions might be proposed for any given problem, each of which is subject to validation. *Problem/solution* in science, then, would indicate that for the given problem "X", "Y" is the most appropriate of all of the possible solutions, and further, that this can be demonstrated by the investigator's experiment. *Problem/solution* means something more like (9).

(9) Because we have found the problem "X", we proposed that the solution "Y" be tested in the manner described in this report.

Problem/solution, then, appears to be a specialized form of the *cause/effect* relationship which serves a function unique to the type of discourse in which it is used, that is, the introduction to an experiment report.

Turning to the other discourse relationship which Figure 3 and Figure 4 have in common, we find the *concession/assertion*, or *true X/yet Y* relationship. It will be easier to discuss this particular part of the Diagram if we have reference to an actual section of discourse in which it appears. Consider, therefore, Figure 5.

Figure 5.

There have been numerous recent reports of the immunologic, biochemical and histochemical properties of the neoplastic cells of hairy cell leukemia. Little attention has been devoted, *HOWEVER*, to the stromal changes in involved organs in this disease. (Nanba, *et al.*, 1977)

Figure 5 indicates that certain information has been reported and that other information has not been reported in a certain area of pathology. This is sometimes called a review of the literature, or a review of previous research. We are concerned with this content, but we are also concerned with the use of the *true X/yet Y* relationship. As Figure 6 indicates, *X/and Y* could have been employed.

Figure 6.

There have been numerous recent reports of the immunologic, biochemical and histochemical properties of the neoplastic cells of hairy cell leukemia. *AND* little attention has been devoted to the stromal changes in involved organs in this disease.

Figure 6, however, strikes us as rather unnatural. Contrast, rather than simple coordination, expresses in an explicit way the nature of problems, in the context of experimental reports. In a scientific sense, a problem is a situation in which some sort of conflict exists. As in Figure 5, the conflict may be between availability of information in one area and the lack of information in another. In other cases, the conflict may be a "dissonance" (cf. Borkin 1978) between two competing theories. The use of the explicit *contrast/contrast* relationship, then, is to be understood as the expression of some sort of conflict within the area of interest. That conflict is the *problem*. Based on the content of reports like Figure 5, the terms *previous research* and *data lacking* have been developed as descriptions for the two units of discourse involved. Note that using the expression *previous research/data lacking* also means *true X/yet Y* will be used explicitly.

Let us turn now to the effect unit of Figure 3. Figure 3 illustrates one of the more common forms of development for the introduction, in which the *effect*, or *solution*, is a single unit. Consider, for example, Figure 7.

Figure 7.

> *The purpose* of this pilot study is to examine the educational environment of the bilingually-schooled child . . . (Bruck and Schultz 1977).

As Figure 7 indicates, this type of introduction contains a statement of the purpose of the experiment. There is no specific hypothesis stated (contrary to the predictions of Lackstrom, et al.), but there is the suppressed prediction that certain kinds of information will be gained. In Figure 7, the information is the educational environment of two children. This type of *solution* unit has been called PURPOSE, to indicate that it is a statement of the purpose of the experiment, although not necessarily an explanation of the hypothesis it is designed to provide evidence for. Although it may be further elaborated, as was the original from which Figure 7 is excerpted, it has a single rhetorical function, one common to all such introductions.

Figure 4 illustrates the other sort of *solution* unit that may appear. Consider Figure 8.

Figure 8.

> With respect to the motor innervation, the question arises whether the alpha and gama efferent fibres would regenerate at different times . . . *it was supposed* that the failure of regeneration of an intact gamma system was responsible for the lack of the tetanus symptomology in these animals.
> IF this interpretation is correct, the marked differential of the motor fibres would provide an interesting new means of obtaining a preparation in which the alpha innervation is still absent. The intent of the present paper is to provide more directly supporting evidence for this. (Nanba *et al.*, 1977)

The first sentence is, in fact, part of the *Problem* unit, being a summary of the first several paragraphs in which the Problem is developed, as we have suggested above. Then, following a brief stretch of discourse in which reasons for a particular solution are cited, the solution is presented, "it was supposed that . . ." The hypothesis is, of course, that there is a difference in the rate of regeneration.

The second paragraph is what we have referred to as the D of C. It consists of a conditional *cause/effect, if X/Y*. If the hypothesis is true, then, the result will be as indicated, "an interesting new means of obtaining a preparation in which the alpha innervation is still absent". Finally, the now-familiar Purpose statement appears.

Based on the evidence of introductions of the sort depicted in Figures 5, 7, and 8, we have developed an idealized version of the Introduction to an Experimental Report. It contains each of the discourse functions noted above, and employs a simple convention to show where the rhetorical functions correlate. Consider Figure 9.

Figure 9.

PROBLEM		SOLUTION (PURPOSE)		
PREVIOUS RESEARCH-CURRENT THEORY	DATA LACKING-CONFLICTING DATA	HYPOTHESIS	DEDUCTION OF THE CONSEQUENCES	
			HYP	RESULTS OF EXPERIMENT

The rhetorical functions of EST discourse have been entered into the Bloc Diagram in the positions corresponding to the explicit discourse functions which correlate with them. For example, the PROBLEM has replaced the *cause*, while SOLUTION has replaced *effect* at the main level of this Diagram. In order to establish clearly that these are implicit rhetorical functions of EST, rather than the explicit discourse functions, they have been printed in upper case letters. Note that PROBLEM/SOLUTION is a specific type of *cause/effect* relationship; the use of the term PREVIOUS RESEARCH/DATA LACKING means that there is a *true X/yet Y* relationship involved, and so on. Although we have not yet dealt with the terms CURRENT THEORY/CONFLICTING DATA, they have also been placed in Figure 9 since many PROBLEMS are developed through this procedure.

Thus far, our use of examples has been limited to isolated sections. To conclude, let us look quickly at one complete section of discourse comprising the introduction to a report published in a recent TESOL Quarterly in Figure 10.

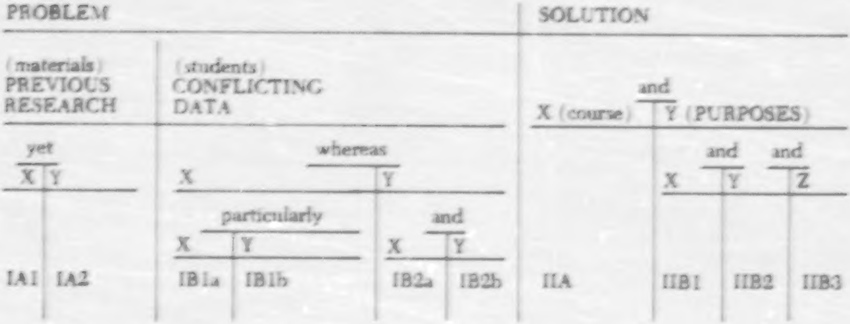
Figure 10.

A Final Sample Introduction

(IA1) Much teaching material has become available recently in the field of English for Science and Technology for the non-native student. (IA2) YET most of this material is geared primarily toward the teaching of reading comprehension. (IB1a) Our experience with foreign students at the Master's level has led us to believe that the greatest strength of many EFL students is in reading technical English, (IB1b) particularly in their field of specialization, (IB2a) WHEREAS their writing skills are seldom as well developed. (IB2b) The ability to write acceptable scientific English becomes a major concern for those students who must successfully complete a thesis as part

> of their degree requirements. (IIA) THUS we have developed a writing program based on recent research in EST in an effort to help our students prepare for the final stage in their degree programs. *The purposes of this paper*, THEN, are: (IIB1) 1) to identify that part of the literature we found most applicable to classroom use; (IIB2) 2) to indicate the steps we took to translate research findings into course syllabus; and (IIB3) 3) to describe the resulting teaching materials. (Weisberg and Buker 1978)

Figure 11.



In Figure 10, the discourse is developed along the lines we have suggested in the idealized Introduction, Figure 9. Note that the overall relationship is PROBLEM/SOLUTION. The break between these two stages comes at the unit marked (IIA), which begins with the bloc signal THUS. The PROBLEM is developed as a contrast between PREVIOUS RESEARCH, units marked (IA), and CONFLICTING DATA, marked (IB). The previous work referred to is the abundance of EST material designed for reading comprehension. The conflicting data is that, in the authors' experience, the students need more writing material and less reading material. As a result of this situation, the solution (a writing program) was developed. Finally, the SOLUTION unit states precisely the three purposes of the paper, which are, basically, to describe that writing program.

We might comment that it is unusual to find an introduction so clearly and concisely developed as this one. One might postulate, therefore, that the authors have been influenced in a positive way by their work with EST discourse.

5. Conclusions

In this paper I have tried to indicate how the explicit discourse functions, common to any written discourse, can be used with specialized functions within a limited universe of discourse such as that of EST, and how these functions must be interpreted, therefore, as implicit in the use of such discourse functions.

Specifically, we found that the introduction to all experimental reports will employ certain standard discourse functions—*cause/effect*, *contrast/contrast*, and, possibly, *conditional cause/effect*. Each of these is understood to have an implicit rhetorical meaning: PROBLEM/SOLUTION for *cause/effect*, contrast between PREVIOUS RESEARCH and LACKING DATA for *contrast/contrast*, and the DEDUCTION OF THE CONSEQUENCES OF THE HYPOTHESIS for *conditional cause/effect*. Although the introduction is not limited to just these functions, and although they may appear in widely divergent forms, we can predict that they will appear and that the development of the Introduction will depend on them.

There is a rhetorical competence involved in the proper interpretation of these implicit rhetorical functions. This rhetorical competence can and should be part of any advanced EST reading or writing course for ESL students.

Part III

Testing Issues

Test Bias in Language Placement Examinations

Hossein Farhady

University of California, Los Angeles

This study investigated possible unfair decisions made on the basis of the results of placement examinations which consist of a variety of subtests. In 1961, Carroll argued against generally accepted discrete-point tests and recommended tests of integrative skills. Since then, there has been an ongoing debate among scholars, some of whom advocate discrete-point and others who prefer integrative tests. However, a question which has not been empirically answered is whether students from different linguistic, educational, and cultural backgrounds are disadvantaged in their performance on either discrete-point or integrative tests.

To address the issue, the performance of foreign students on these two types of tests was investigated. The subjects were 350 incoming foreign students representing eight different countries. The instrument used was the placement examination at UCLA which consisted of eleven subtests.

The results of this study strongly support the hypothesis that some foreign student groups perform better on discrete-point subtests and others on integrative subtests. Finally, the effects of these differences on placing foreign students in appropriate ESL courses are discussed and three alternatives are suggested to avoid disadvantaging any group of students.

The great number of foreign students attending American universities has intensified the problem of placing these students in appropriate ESL courses. In most cases, placement examinations consist of a variety of subtests whose purpose is to assess student competencies in various language skills (listening, reading, writing, and speaking) and in various linguistic areas such as phonology, morphology, and syntax. Often, incoming foreign students are not equally competent in all language skills, due to their educational, cultural, and linguistic backgrounds. This diversity may influence their performance on language proficiency tests used for placement purposes. This paper attempts to demonstrate the effects of different types of tests in assigning foreign students to various language courses as well as possible unfair decisions made on the basis of test scores.

1. Background

Teaching and testing are sometimes appropriately referred to as a single process because modifications on the former have influenced the latter. It was not until the rise in prominance of the audio-lingual pedagogical approach that the ESL teaching strategies, and consequently testing procedures, underwent fundamental changes. Thereafter, the theory of language testing developed by Lado (1961) and others resulted in the so-called *discrete-point* testing approach.

The advocates of this approach believe that, since language is comprised of sounds, intonation patterns, morphemes, and the meaningful arrangement of words, testing these elements would enable teachers to effectively assess their students' abilities in various language skills.

Later, when the audio-lingual method was challenged by proponents of other approaches, the testing procedures developed under the influence of the audiolingual method were also subject to question. Carroll (1961) was the first to trigger the reform by stating, ". . . I recommend tests in which there is less attention paid to specific structure points or lexicon than to the total communicative effect of an utterance" (p. 318). Carroll referred to this type of testing as the *integrative approach*. Scholars are still debating the choices, some advocating discrete-point tests, others tests of integrative skills.

Much has been written on the adequacies and shortcomings of both discrete-point and integrative tests (Oller 1973a, 1974, 1976b; Brière 1970, 1972; Rand 1972; Jakobovits 1970; Spolsky 1973; Hinofotis 1976). However, an uninvestigated issue is that students from different linguistic and educational backgrounds may perform differently on discrete-point and integrative tests.

The purpose of this paper, then, is to investigate: 1) potential differences between the performance of foreign students on discrete-point and integrative tests, 2) the effect of the sex variable on the subjects' performance on discrete-point and integrative tests, 3) possible unfair decisions made due to biased test scores, and 4) the effect of selecting one test over the other for student placement in ESL courses. The final section of the paper is devoted to suggesting alternatives for overcoming some of the problems with existing placement procedures when placement is based on the composite score of several subtests.

Various Subtests of the ESLPE

Discrete-Point		Integrative	
Spelling	22 items	Reading	23 items
Punctuation	22 items	Listening	30 items
Structure	62 items	Dictation	50 items
Vocabulary	45 items	Cloze	50 items

2. Method

2.1 Subjects. The subjects in this study were 491 incoming foreign students who took the Fall 77 version of the UCLA English as a Second Language Placement Examination (ESLPE). They represented eight linguistically and educationally different groups. Table 1 represents the number of subjects from each country.

Previous research on the ESLPE (Sanneh, 1977) demonstrated that factors such as major field of study and academic status did not have any significant effect on the performance of foreign students. Consequently, the subjects in this study were not categorized on the basis of the above mentioned variables.

2.2 Instrumentation and Procedures. Since 1956, the University of California, Los Angeles has been administering different versions of the ESLPE to all incoming foreign students whose native language is not English. Research on

TABLE 1

The number of students from different countries taking the Fall-77 ESLPE

Country	Number of Students
Iran	56
Taiwan	51
Korea	53
Japan	44
Hong Kong	27
France	27
Mexico	25
Israel	22
Others	186
Total	491

various forms of the ESLPE indicates that the composite score of the subtests tends to accurately measure the students' overall language proficiency (Connolly 1972). The Fall 77 version of the ESLPE consisted of eleven subtests, evenly divided into two subsections. The discrete-point subsection consisted of spelling (22 items), punctuation (22 items), structure including articles, prepositions, and verb forms (62 items), and vocabulary both Romance and Germanic (45 items). The integrative subsection included reading comprehension (23 items), listening comprehension both visual and written (30 items), dictation (50 items), and cloze (50 items). The subjects taking the test were allowed three and a half hours to complete the test and a bio-data sheet. The ESLPE scoring system allowed one point for each correct response. The students were not penalized for guessing, and blanks were counted as wrong answers. The cloze test was scored by the exact-word scoring procedure.

It should be mentioned that classifying a test as either discrete-point or integrative seems to be difficult and sometimes inappropriate. Most testing specialists, including Oller, believe that the difference between discrete-point and integrative tests is a matter of degree rather than type. We may consider the two along a continuum ranging from highly discrete-point items at the one end to highly integrative items at the other (Oller 1973b). Nonetheless, discrete-point tests generally aim at testing one and only one point of language at a time. Integrative tests, on the other hand, are intended to assess language proficiency in terms of holistic performance, sometimes akin to actual functions of the language in real life situations. However, due to their nature, the subtests of the ESLPE were easily categorized as belonging to either the discrete-point end of the continuum or the integrative end.

2.3 Data Analysis and Results. Subtest scores were standardized to achieve equal weights before adding the total score. Then two separate analyses were conducted.

Analysis 1: The standardized scores of the 305 subjects were analyzed utilizing a two way analysis of variance (ANOVA). The results of the ANOVA, presented in table 2, supported the hypothesis that some foreign student groups, depending on their educational and linguistic backgrounds, performed better

TABLE 2

The scores of eight language groups on DP and IN subsections*

Country	DP		IN	
	Mean	SD	Mean	SD
Hong Kong	218	27.15	221	24.84
Israel	201	28.70	218	29.98
France	204	32.59	213	34.62
Iran	188	35.82	198	34.34
Mexico	199	35.55	199	39.10
Taiwan	209	38.34	200	27.70
Korea	196	33.68	187	31.52
Japan	192	35.37	184	33.80

* DP. Discrete-point
IN. Integrative

on the discrete-point subsection and others performed better on the integrative subsection.

It is clear that, except for Mexican students, who scored equally well on both types of tests, there were significant differences among students from various countries. Students from Iran, France, Israel, and Hong Kong scored higher on the integrative subsection. Students from Taiwan, Korea, and Japan scored higher on the discrete-point subsection. The results of this analysis also indicated that students from Hong Kong and Japan obtained the highest and the lowest total scores on the ESLPE, respectively. Finally and more interestingly, the difference between the scores on the discrete-point and integrative subsections was greater for the students from Taiwan and Israel than for students from other countries. However, the difference was in the opposite direction, i.e., students from Taiwan scored significantly higher on the discrete-subsection, whereas Israeli students scored higher on the integrative subsection. Figure 1 illustrates the relative standing of the students from each country.

Analysis 2: In this analysis, the scores of male and female students (N= 491) on the discrete-point and integrative subsections of the same ESLPE were compared. The data were analyzed using a regression model of discrete-point subsection (Y) on the integrative subsection (X_1), sex (X_2), and the interaction between the two (X_1 and X_2). The results of this analysis, which are reported in table 3, indicated that eighty percent of the variation in the discrete-point subsection was accounted for by the integrative subsection. The contribution

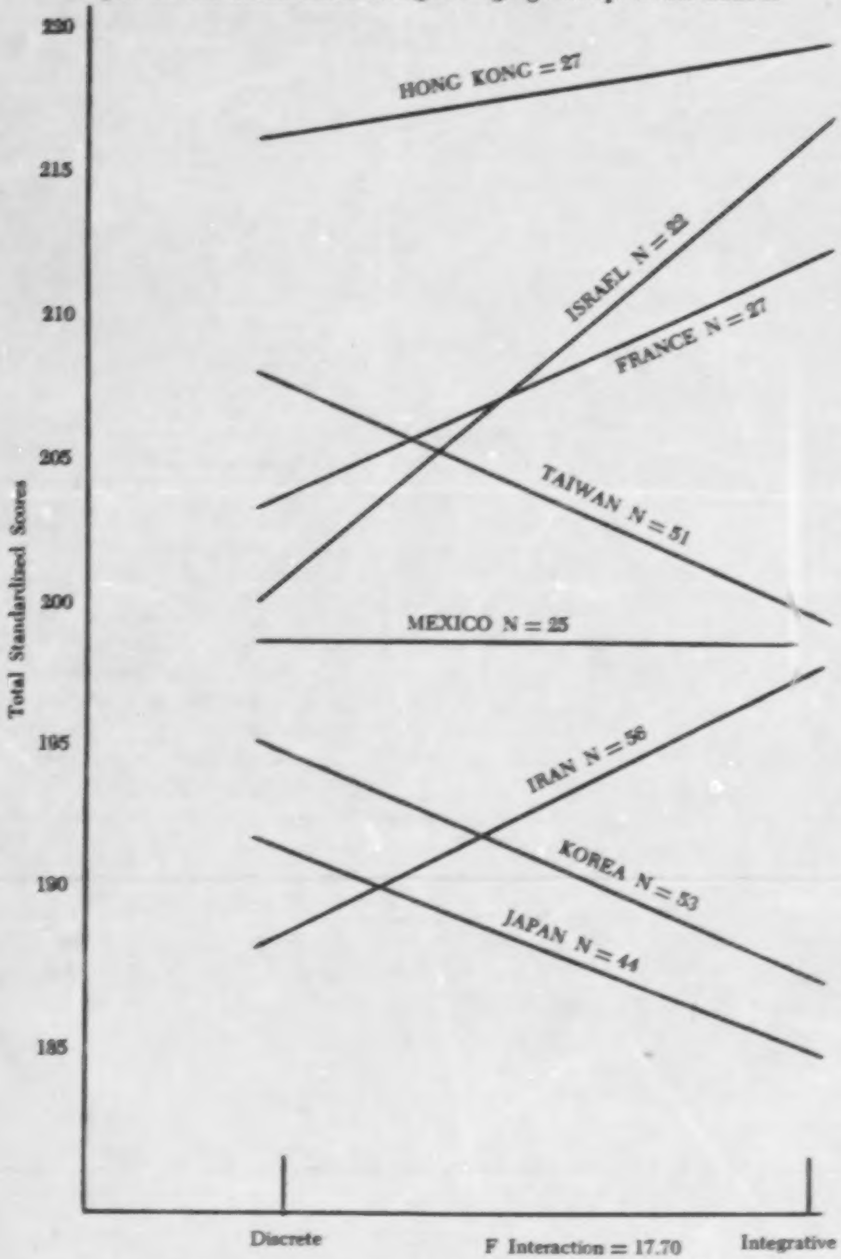
TABLE 3

Regression Statistics on DP and IN

Variable	MR	R^2	b	SEE	F	Mean	SD
Integrative	.89	.80	.85	11.41	1990**	91.52	25.68
Sex	.89	.80	9.97	11.38	7**	.67	.47
Sex X Integrat	.89	.80	−.10	11.33	6**	61.59	49.58
Constant = 23.40							

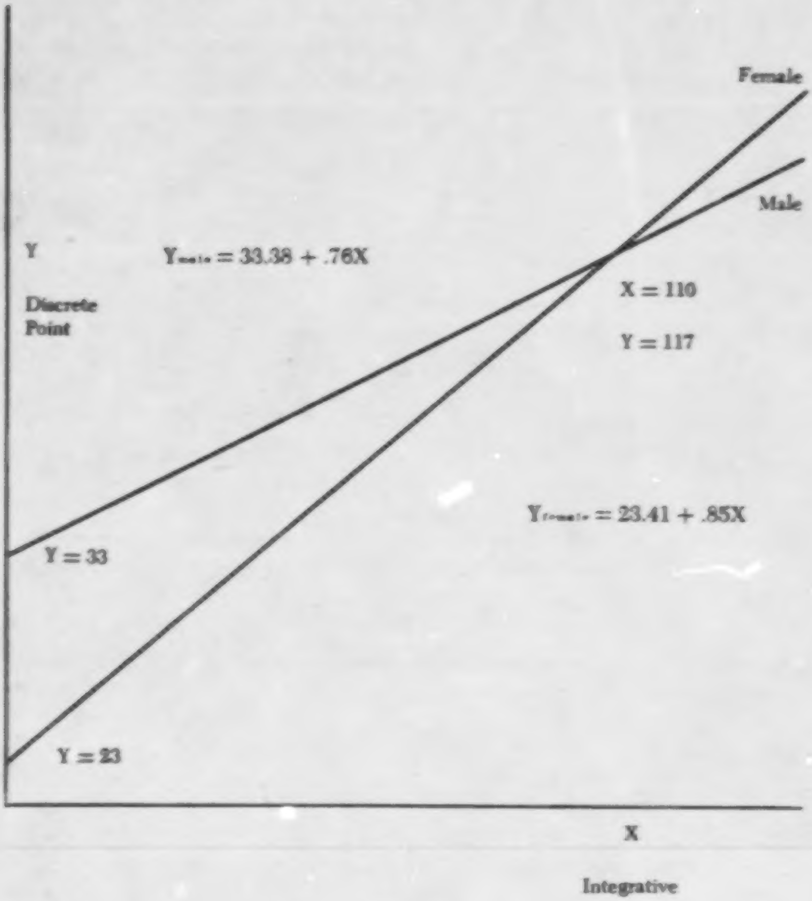
** $P < .01$

Figure 1. The Performances of Eight Language Groups on the ESLPE.



F Interaction = 17.70
Significant at .001 level

Figure 2. Regression Lines for Males and Females.



of the sex factor and its interaction with the integrative subsection added nothing to our information when included in the model simultaneously.

The negative value of the regression coefficient for the interaction effect indicated that the two regression lines intersect at $X=110.70$ and $Y=117.57$. Thus, below this point, males outperformed females on the discrete-point subsection, whereas above the point of intersection, females scored higher than males. This interaction effect is illustrated in figure 2.

To establish regions of significance, the formula developed by Johnson and Neyman (1936) was utilized, with the range of significance calculated to be between X=136.20 and X=94.80. This implied that, within the range of these values, the scores on the discrete-point subsection were not significantly different across the groups. Nonetheless, students who scored beyond and below these values on the integrative subsection scored significantly different on the discrete-point subsection. Since the possible attainable scores on the discrete-point and integrative subsections were well below and beyond the range of significance (151), it seemed important to consider sex as a potential variable which could influence the performance of foreign students on placement examinations. Since the results of this analysis did not confirm the findings of previous research mentioned above (Sanneh 1977), until further investigation, these conclusions should be considered speculative rather than definitive.

3. Discussion

The notions of discrete-point and integrative tests deserve serious attention in the development of second language proficiency tests. In this study, students from different sexes as well as different educational backgrounds were found to perform better or worse on one or the other type of test. No matter how reliable, valid, and practical a given test may be, there will be some undesirable biases due to variables which are not directly relevant to the underlying constructs of the tests.

The UCLA ESLPE, a statistically well established instrument, having a high internal consistency (.97) and being validated against several criteria, could be considered a good placement test. However, depending on the number and/or weights of the subtests, a student from a specific educational background could be placed in either a higher or lower level course than is justified. An example may help clarify the point. Suppose that there are two types of tests (test A which is discrete-point and test B which is integrative), two groups of students (group 1 from Taiwan and group 2 from Israel, which both have similar proficiency levels), and three levels of ESL courses (elementary, intermediate, and advanced). If the two groups took test A, the students from Taiwan would perform better than the students from Israel. Thus, by obtaining high total scores, they would be placed in the advanced course; the Israeli students would be placed in the elementary course because they would not be able to perform to their fullest potential on a discrete-point test (see Figure 3). On the other hand, if test B was administered to the same groups, the reverse would occur (see Figure 4).

There is still a third possibility wherein the two tests are combined and given to the same groups. In this case, both groups would obtain fairly equal total scores and be placed in the same level (see Figure 5).

Such hypothetical placement may seem exaggerated, but it could occur. In short, regardless of the abilities of the examinees, some will be favored or penalized because of the nature of test being administered. Keeping the above discussion in mind, the following points seem important.

1. As long as the controversy between discrete-point and integrative tests

Figure 3. Hypothetical Placement of Students

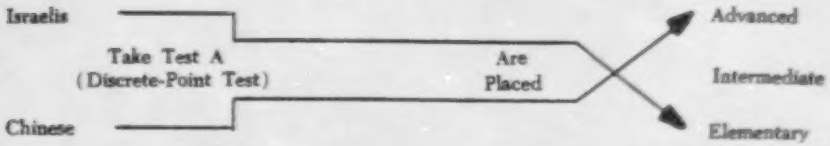

Israelis
Take Test A
(Discrete-Point Test)
Chinese
Are
Placed
Advanced
Intermediate
Elementary

FIGURE 4

Hypothetical Placement of Students

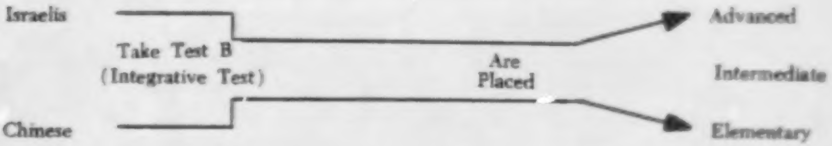
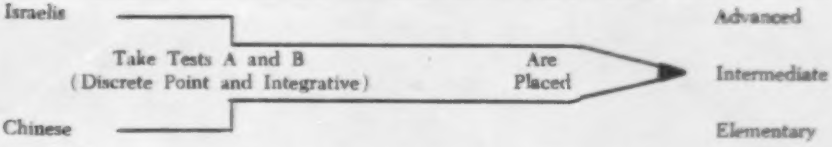

Israelis
Take Test B
(Integrative Test)
Chinese
Are
Placed
Advanced
Intermediate
Elementary

FIGURE 5

Hypothetical Placement of Students


Israelis
Take Tests A and B
(Discrete Point and Integrative)
Chinese
Are
Placed
Advanced
Intermediate
Elementary

is not resolved empirically and definitively, attempts should be made to keep a balance between the two in language proficiency tests. Since the underlying constructs of overall language proficiency have not been identified, it seems advisable not to assess the language proficiency of the students with only one specific type of test.

2. Since ESL students in academic environments need different language skills than their non-academic counterparts, it seems important to distinguish the extent to which different language skills (survival and academic) are required of university students in competition with native speakers. Of course, students are required to use both types of language skills, but academic English is more important to them than to a tourist or a person who intends to live in an English speaking community. Students are required to listen to lectures, read textbooks, participate in class discussion, be familiar with spelling and mechanics of writing, and produce acceptable written and oral forms of language; such a demand is not required of a non-student foreigner. Therefore, university proficiency tests, especially placement examinations, should be more thorough than other types of tests by including academically important as well as survival language skills.

3. In the case of course placement, as mentioned before, the procedure for using the total scores as a basis for placing students in different language courses has flaws. The main limitation of this procedure may be resulting differences in the linguistic and communicative abilities of the students placed at the same level of instruction. A student who is proficient in grammar is not necessarily competent in communicative skills. Also, a student who feels comfortable in communicating with native speakers is not necessarily highly skilled in utilizing discrete linguistic elements of the language. Therefore, instead of placing students on the basis of their total scores, they may be more appropriately placed according to their individual language abilities, which may be reflected in their performance on the discrete-point and integrative tests. As a result of such a placement procedure, instead of having sequential courses, skill-focused classes might be more appropriate. In this case, a student who was quite competent in one skill would not be required to take the course which emphasizes that skill. Instead, he would take courses where the emphasis was on the skills necessary to better fulfill his language needs.

It is true that language skills are highly interrelated, but still there seems to be an imbalance in the abilities of foreign students in different skills. This may be due in part to the absolute difficulties of the skills as well as to the previous education of the students in which certain skills received more emphasis than others.

In summary, the conclusions drawn from the results of this investigation and their implications are important for placement purposes, especially in the United States. The suggestions will hopefully encourage officials and test developers to pay close attention to the type of the test as well as the language and educational backgrounds of the examinees in order to avoid unfair decisions made on the basis of test results.

Cultural Bias in Reading Comprehension Tests

Bernard Mohan
University of British Columbia

Some test items in reading comprehension tests test both language knowledge and cultural knowledge. Various standardised tests contain them and thereby discriminate against ESL students. This paper demonstrates how test items for commonly used reading comprehension tests are culturally biased, indicates the degree and location of bias in a number of popular tests, and outlines a procedure for detecting bias that teachers can use. The view is taken that evaluating these test items is an application of linguistic semantics, which means that judging their 'content validity' can be put on a solid theoretical basis.

Reading comprehension is one of the various activities that ESL students perform where differences of culture can lead to differences of interpretation. In test situations such differences of interpretation become serious difficulties. This paper will examine the problem that the culturally different pose for reading comprehension tests. Three questions are of interest:

a) Is there cultural bias in reading comprehension tests, and if so, does it matter that there is bias (e.g., does the bias amount to discrimination)?
b) How can teachers and students, the consumers of tests, identify bias?
c) Is cultural bias in reading comprehension tests inevitable in principle, or are there systematic ways in which it can be avoided or at least minimized?

These questions will be approached through the analysis of question items from tests. However the analysis will be linguistic rather than statistical and a particular aim will be to show how teachers can do such an analysis for themselves.

Consider the following test items where a student will have to rely on specific socio-cultural knowledge to get the right answer. The items come from widely used tests of reading comprehension: the Stanford Diagnostic Reading Test, the Canadian Test of Basic Skills (a revision of the Iowa Test of Basic Skills for Canadian use), the Gates McGinitie Reading Test, the Nelson-Denny Reading Test and the Gapadol Reading Comprehension Test. They have been chosen to show a range of cultural knowledge and have been altered non-significantly for reasons of copyright and brevity. (In each case the form of the test and the item number is given so that the original can be identified).

1) *Patriotic objects:* 'There are red & white stripes and white stars in our flag. Our flag contains one __________ for every state.'
a) stripe b) star (S.D.R.T. Form X. Level 1. #32)

2) *Food:* 'In the story the French regarded potatoes like most Canadians regard:'
a) spinach b) tomatoes c) horsemeat d) margarine
(C.T.B.S. Form 1, #89) (Note: in the story the French dislike potatoes.)

3) *Customs:* 'In this poem, what does "April Fool!" mean?'
(C.T.B.S. Form 1, #98)
a) The person who said it was fooling b) It was not April at all.

4) *Games:* 'Sam won at marbles because he could __________ straighter than Bill.'
a) show b) shoot c) draw d) run
(G. McG.D. Form 3M. #1)

5) *History:* 'In the first colonies in America, making clothing took time. The women first had to spin the yarn. Clothes for the colonial family were usually made in __________.'
a) factories b) homes c) luxury d) China
(G. McG.C. Form 1. #14)

6) *Geography:* 'The Yankee peddler traded as far west as the Mississippi and as far south as Louisiana. He operated __________.'
a) over most of the country b) as far south as Louisiana.
(N.D. Form D. #10)

7) *Folklore:* 'Pam went to the party with a tall pointed black hat, long black cape and a broom. She was dressed as a __________.'
a) witch b) ghost c) cowgirl d) pumpkin
(G. McG.C. Form 1. #4)

8) *Housing:* 'Bill ran out on his front porch to watch the firetruck. He lives in __________.'
a) a big apartment b) a city house c) a trailer
(C.T.B.S. Form 4. Level 9 #1)

9) *Culture-bound metaphors:* 'What does "applying soft soap" mean in paragraph 6?'
a) looking clean in public b) flattering people (C.T.B.S. Form 1. #131)

10) *Stories familiar to particular cultural groups:* (Cloze test example)
'The person was holding tight to the handle of an open __________, dangling by one hand like a doll tied to the string of a balloon . . .'
(G. Form Y. #2)

In order to answer these one needs to know about (successively): the symbolism of the U.S. flag; Canadian food preferences; April 1st customs; how marbles is played; early American history; American geography; the garb of the typical witch; the layout of different types of North American homes; an English idiom; and the story of Mary Poppins. It is clear that special cultural knowledge is required, and therefore that there is indeed cultural bias in reading comprehension texts.

The procedure used to identify biased items can be summarized in two questions:

11) Does the item test language knowledge, or does it test knowledge of the world too?

In (12) it might seem that knowledge of the world concerning witches and Halloween is needed to answer the item. Yet on closer inspection it becomes clear that the item can be answered on the basis of language knowledge since answer (b) is the only one which fills the slot to make a semantically acceptable sentence. The second question:

12) The playmates wore costumes to Sandra's Halloween party. Nellie __________ a tall, black, pointed hat.
a) walked b) wore c) cared d) hurt e) looked
(G. McG. D. Form 3M. #7)

13) Does the item test: a) knowledge of the world which is available to all cultures or b) knowledge of the world which is readily available only to particular cultural groups?

Item (14) tests knowledge of the world. This, not language knowledge, is the basis for seeing a connection between playing all day and feeling tired. However, since it is reasonable to assume that every cultural group would see such a connection, the item does not depend on a piece of special cultural information. It does not test knowledge of the world available only to particular cultural groups.

14) Sam and Bill played the whole day. In the evening they felt __________
a) rested b) small c) tired
(S.D.R.T. Form X, Level 1, #17)

By this procedure items (12) and (14) are not considered to be culturally biased, although they might appear to be. On the other hand, items (1-10) are considered culturally biased, for in each case special cultural information is required and the answer to question (13b) is positive.

Having established that cultural bias exists, the next question is whether the presence of bias is an important matter. Certainly at least three of the tests from which items (1-10) are drawn are widely used: the Stanford Diagnostic Reading Test, the Gates-McGinitie and the Canadian Test of Basic Skills. One aspect of this question is the frequency and distribution of bias. For example, one can count the percentage of biased items in a test. My investigation of the Canadian Test of Basic Skills, Form 1 (grades 3-8) indicates that 7% of the items are biased. This is by no means negligible in its likely effect on the scores of culturally different students. Moreover the biased items are unevenly distributed, so that in the grade 5 section of the test the percentage of biased items is 15%. Other tests show such clustering too. Thus the Gates-McGinitie Primary C Form 1 test contains 10% biased items and in the Stanford Diagnostic Reading Test Form X Level 1 the percentage rises to 16%. These percentages are high enough to have a marked effect on scores. In addition, these figures are somewhat conservative and almost certainly underestimate the amount of bias. Items which are borderline cases have not been counted, and my analysis of the items, as a member of the host culture, is likely to overlook some of the less obvious cases of bias.

Another aspect of the importance of bias is the question of whether the decisions made about individual students on the basis of biased reading comprehension tests are trivial or serious. For example, reading comprehension scores are often used to assign students to reading groups of different levels of competence. Here a biased result is likely to lead to incorrect diagnosis of difficulties in reading and inappropriate teaching. More serious is the situation in some schools where reading comprehension scores are the means to place students in streams of different general academic ability i.e. a low score puts students in slow class, possibly for the rest of their school career. When a reading comprehension score is used as an index of academic ability, a biased score and the decision following from it can have all the force of a self-fulfilling prophecy.

Even if it is accepted that bias exists in quantity and that it may affect important decisions, it is still possible to argue that bias is not a matter worthy of serious attention. One argument is that students should be tested on cultural knowledge since they will not be able to function properly in North American society if they do not learn about North American culture. A reply to this is that it is certainly appropriate to test for cultural knowledge but that testing for cultural knowledge should not be confounded with testing for reading comprehension. If one tests for cultural knowledge one should test for it separately, specifically and systematically, as some Modern Language teachers do. A further and much stronger argument denies that cultural knowledge should be tested separately and asserts that special cultural knowledge should be a necessary part of reading comprehension tests. The argument (which has been advanced by competent and thoughtful ESL teachers) is based on the claims that special cultural knowledge will inevitably occur in reading materials that students will have to deal with, and that a reading test should faithfully reflect the reading tasks and materials facing the student.

While there are a number of objections to be made here (e.g., how much special cultural knowledge is actually required by a math textbook?) what is particularly wrong with this view is that it concentrates on the students' performance but ignores the evaluation being made of that performance. Suppose two students give the wrong answer to a reading comprehension item. One is Canadian, the other a recent immigrant. Both have misunderstood, but the evaluation made of the first may not hold for the second. The former may be weak in reading comprehension, the latter may be missing some cultural knowledge. The task they both face is the same, but the interpretation of what they do is different. One cannot automatically conclude that the immigrant is weak in reading comprehension.

From what has been said so far, various suggestions can be made about the second question: how teachers and students can identify bias and what action they might take. First, teachers can investigate the items in the reading comprehension test which is currently being used in their school and in any other available reading test which might be used in its place. A procedure for doing so has been outlined above in (11-14). Group discussion of the items, either with other teachers or with ESL students, works well since the lone reader often passes over difficulties that a group will detect and makes cultural assumptions that ESL students will be more conscious of, and it is an activity which is justifiable in itself, since it requires careful reading and interpretation. If the current reading comprehension test turns out to be more biased than an available alternative, then a case can be made for switching to the alternative test. Secondly, once biased items have been noted, teachers can check to see that the bias is taken into account in the interpretation of test scores and in the decisions about students that are based on such interpretation.

Finally we turn to the third question: whether cultural bias in reading comprehension tests is inevitable in principle, or is avoidable in systematic ways.

What has been demonstrated so far is that some test items confound language knowledge and cultural experience. In effect these items are not simply testing the ability to read the comprehension passage, they are testing for other information as well. As such, they are poor tests of reading comprehension. The maker of standardized tests has certain procedures of statistical item-analysis which are intended to ensure that items meet standards of validity and reliability, yet these do not seem to have been successful in eliminating culturally biased items in the tests examined. The reason for the presence of biased items appears to be that the initial scrutiny of items for suitability (or content validity) is very crude and allows biased items in. The remedy offered here is to relate the analysis of items (as shown in 1-14 above) to semantic theory in order to show that the analysis of items can be based on semantic principles. I believe that this is helpful for clarifying both cultural bias and the general issue of the content validity of reading comprehension items.

The heart of semantic theory is the relation of inference (or consequence of implication). If one can show the role of inference in reading comprehension items, one can show the connection between the analysis of these items and semantic theory. (15) is a clear example of the inference/implication/consequence relation. "The girls were putting icing on the cakes" Passage.

15) ". . . Three girls put frosting on the cakes that had cooled. We learned a new word for "frosting". At the Bakery it was called "icing" . . ."
Q: What were the three girls doing?
A: 1) cooling the cakes
2) *putting icing on the cakes*
3) making frosting for the cakes

(C.B.T.S. Form 1, #9)

Can the answer to 15 be inferred from the passage? The passage implies that answer. The passage does not imply answer (1) and it does not imply answer (3). The general form of the relations between the passage and the answer is given in (16a), and this general form can be simplified still further, as shown in (16b) to the question of whether one sentence (S_1) implies another sentence (S_2). Thus (15) at its simplest requires the reader to decide whether "three girls put frosting on the cakes that had cooled" implies "the girls were putting icing on the cakes". (A check on this is that "therefore" or "consequently" can be appropriately inserted between these two sentences but not between the passage sentence and the other possible answers). This is the typical pattern for reading comprehension test: the correct answer sentence can be inferred from one passage sentence (or a combination of passage sentences), and more generally, the student is being asked to judge whether one sentence follows from another or not. In this way inference plays a central role in reading comprehension items.

16) a) Passage implies right answer sentence
Passage doesn't imply (or contradicts) wrong answer sentence
b) S_1 implies S_2 — true? false?

There are a number of variant forms of the pattern shown in (16a). For example, (14) takes a different form, in that most of the answer sentence is part of the passage. Yet when the correct word is inserted the second sentence follows from the first. "Sam and Bill played the whole day" implies "In the evening they felt tired." It does not imply the possible sentence made by the insertions of "rested" or "small." (14) conforms to the general form of (16b). However, this is not to say that every comprehension question conforms to (16b). Cloze items like (10) minimally require that the word inserted in the slot should complete the sentence to form a semantically acceptable sentence. If that is all that is required in an item, it is demanding recognition of relations within a sentence rather than recognition of implication, which is a relation between sentences.

The concept of implication has an important bearing on the analysis of cultural bias. It has already been noted that the cultural bias problem in reading comprehension tests arises because both language knowledge and cultural knowledge are required in these tests. Clearly it would be helpful to distinguish between language knowledge and cultural knowledge in a principled way. (15) is an item which tests language knowledge and (7) is one which tests cultural knowledge. Both require the recognition of an implication. The implication in (15) is based on language knowledge and the implication in (7) is based on cultural knowledge. The distinction between these two kinds of implication is almost identical to the distinction semanticists like Katz and Leech have made between semantic implication, which is based on knowledge of language and factual implication, which is based on knowledge of the world (which would include cultural knowledge).

This is a very important and useful connection. Even though it is not uncontroversial and not without borderline problems, the distinction between semantic implication and factual implication is based on clearly argued principles (Katz 1972, pg. 117ff; Leech 1974, pg. 7ff) and the same arguments can be used to establish the distinction between language knowledge and cultural knowledge in test items. In addition, a major achievement of linguistic semantics is the detailed analysis of semantic implication in natural language. This means that the analysis of semantic implication in reading comprehension test items can be supported by a developed body of knowledge. This is not to say that the body of knowledge is complete and the analysis is automatic for a great deal still needs to be done. Yet it is clear that the analysis can be systematic.

Nor is this to say that reading comprehension test items should be restricted to those that solely test semantic implication. (17) identifies a number of different bases for implication. (17b) requires some further explanation. In (15) the passage tells the reader that "frosting" and "icing" are to be considered as

17) *Implies* on the basis of:
a) knowledge of language —Language tests
b) information in the passage (gained by using (a))
—Reading Comp. tests e.g. (15)

c) general knowledge of the world (supporting (a) & (b))
—Reading Comp. tests e.g. (14)

d) special cultural knowledge —Cultural tests e.g. (7)
e) special 'subject' knowledge —Achievement tests.
(a) = Semantic/Linguistic implication
(c), (d) & (e) = factual implication.

synonyms, and skilled readers will treat the words as such even if they are not synonyms in the readers' own usage. In other words, information in the passage is capable of overriding the speakers' personal knowledge of their language. This is an extreme case to illustrate the importance of passage information. The use of passage information is a necessary feature of reading comprehension test items, for otherwise it would be possible to answer the items without reference to the passage. Since (13) implies that general information about the world (17c) is also acceptable, it is claimed here that 17 (a-c) are permissible in reading comprehension test items. It has been argued above that special cultural knowledge (17d) is not permissible and 17(e) indicates that the same holds for the type of special subject knowledge that is typically included in achievement test items. The distinction between semantic and factual implication would draw the line in a different and inappropriate place, excluding every thing but (17a). So semantic theory is helpful because it provides a principled basis for making distinctions which are essential, but reading comprehension test items are not limited to semantic implication and reading comprehension tests are not equivalent to semantic tests. A final point is that the exclusion of (17d) and (17e) is meant to hold for tests only, and not for classwork in reading comprehension. In a test the lack of special cultural information can be a stumbling block whereas in a classroom situation it can be a valuable point of discussion.

To summarize, items like (1-10) are illegitimate in tests of reading comprehension when taken by ESL students because they require special cultural knowledge which is available only to particular cultural groups. This special knowledge need not be incorporated into reading tests. Reading tests should essentially test the ability to use knowledge of the language and reading skills to get meaning from the printed page; it is the task of achievement tests (of biology, social studies, etc.) to test for non-linguistic knowledge. It is legitimate to draw on non-linguistic knowledge in a reading comprehension test only where it plays a neutral, supporting role, and is common to all students taking the test. It should be standard policy to eliminate items like (1-10) from tests when those taking the test include culturally different groups. In fact this would be likely to improve the test for all groups by eliminating material which is irrelevant to the aims of the test. This elimination can be made using the intuitive procedure outlined in (11-14) above. Linguistic semantic theory plays a crucial role in establishing the major distinctions required and in providing the detailed semantic description necessary to put the procedure on a systematic basis.

Evaluating Bilingual Competence: An Experimental Innovative Test

J. Donald Bowen
University of California, Los Angeles

Sandra Plann
Northrop University

Evaluation of linguistic competence, inherent in language teaching, is particularly important in bilingual education, where a need exists to measure achievement in both languages. In fact testing is mandated where public money is provided, since funding agencies rightly feel there should be a determination of the value of programs supported.

Since both languages are tested, a comparison of proficiency will be more meaningful if we are confident that test results across language boundaries are indeed comparable. Attempts to design equivalent evaluations have led to experimental applications of cloze testing, bilingual syntax measures, etc. The effort should continue: the need is great and even negative reports can be useful.

Encouraged by the success of an experimental integrative grammar test in English (IGT), we wish to apply to Spanish the relatively simple concept of identifying the second word in sentences where phonological features are obscured at word boundaries. The test has proved useful in English, with remarkably high coefficients of validity and reliability. This 100-item test, administered by tape recording in 18 minutes, is very efficient.

This article reports on a project which had its genesis in the development of an English proficiency test. This experimental test attempts to measure a subject's competence in the structure of informal spoken English, employing an exclusively integrative format. The test, referred to as the IGT (for Integrative Grammar Test) rests on a very simple assumption: that ability to handle the morphophonemic alternations in English, especially those that characterize variation at word boundaries in informal oral communication, will correlate meaningfully with competence in using the grammatical patterns of the language. Native-speakers of English do very well on this test, usually scoring 80 to 100 percent, but non-native subjects, particularly those with indifferent control of the morphophonemic features, find the test difficult. The scores of *non*-native speakers range from nil to native-speaker range, and there is evidence that the spread of scores for non-natives reflects with reasonable accuracy the varying levels of ability of the examinees.

The test has been reported on at various conferences and in articles that have appeared in professional journals. We will therefore illustrate the test with just one triad of examples. The task is to listen carefully to a short sentence, determine the full form of the second word, and write it down on a test answer sheet. The subject tries to make the identifications as s/he hears the item sentences:

1. Whattaya want with that cat?
2. Whattaya doin' with that cat?
3. Whattaya done with that cat?

The answers, so easy for a native speaker that s/he is likely to miss a few items because attention wanders, can be surprisingly difficult for non-native subjects. Herein lies the value of the test.

First, let's look at the example items. The answers are *do*, *are*, *have*, but this information is not physically present to the subject, who hears for all three: /wə́təyə/; the /wət/ initial syllable is *what* in each case; the third syllable is /yə/ for *you*. The second syllable, crucial for the test, is reduced to the ultimate —a simple, weak-stressed schwa, pronounced /ə̆/. The examinee has to look further into the sentence for the evidence that disambiguates this schwa, by sorting out the lexical-grammatical co-occurrence potential of each sentence: *want, doing, done* require the interpretations *do want, are doing, have done.*

Indications are that the English version of this experimental instrument is reasonably successful. An administration to 686 applicants for admission to the American University in Cairo developed the following figures for the English version of the IGT:

Validity:	A coefficient of correlation of .871 between the IGT and the grammar subtest of the MTELP (Michigan Test of English Language Proficiency).
Reliability:	A coefficient of correlation of .968 from a test/retest set of scores, inherent in the test design.
Efficiency:	One reading only of each item with an item flow of approximately one every 6 or 7 seconds. The entire test, including instructions and sample items, can be completed in about 18 minutes.
Administration:	The test is presented on tape which helps guarantee comparable conditions for different administrations. The taped delivery also controls the pace of the test, to provide the time pressure that helps keep validity and reliability high.
Correction:	Items are highly objective and can be accurately hand corrected (depending on score and on the clarity of the subject's handwriting) at the rate of one and a half minutes to 3 or 4 minutes per test paper.

Our task, then, was to construct an aural Spanish test which, like the English test, would measure competence in the structure of Spoken Spanish. We began by thinking of items, often in pairs, whose comprehension depends on catching a phonetic, syntactic, or semantic clue. For instance, the clue was phonetic in the pair,

Una perra es lo que quiere esa muchacha rubia

and

Una perra es lo que quiere esa muchacha rubia.

The clue was syntactic in:

Las había memorizado para el examen

and

Las sabía de memoria para el examen,

and, a semantic clue allowed the differentiation between:

Los hirvió por 10 minutos para tenerlos bien cocidos

and

Lo sirvió y después empezamos a comerlo.

To be a valid item, native speakers of Spanish have to be able to identify and differentiate both sentences consistently when presented with the two together, or either one alone. If *they* can't, there's no point in asking students to do so.

We tried these items out with native speakers from various Spanish-speaking countries, and eliminated items that were judged to be too obscure, or to contain regional vocabulary. We tried to be particularly careful in choosing common vocabulary items, to make the test intelligible to both a variety of native speakers and to students of Spanish whose primary contact with the language may have been in the classroom.

After this initial screening of items, we piloted a preliminary version of the test on native speakers. We began refining the exam by then eliminating or modifying the items that everyone, or nearly everyone, failed to identify. For instance, an item in an early version of the test was:

Las elecciones de ayer nos sorprendieron.

Almost everyone wrote *lecciones* instead of *elecciones* for the second word, due, we assumed, to the academic setting in which the test was taken. To clarify this item, we added a semantic clue:

Las elecciones de ayer sorprendieron a los socialistas. Notice that the addition of this bit of contextual information weights the clues heavily toward the semantic component, and a knowledge of Spanish vocabulary will aid considerably in the identification of the second word in the sentence.

On the basis of these preliminaries, we put the fifty items in pairs and ordered them, from the easiest to the most difficult. On the second part of the test, the same fifty items were listed in random order. We then made what we hoped would be a final recording of the examination. It was recorded by a speaker of Castilian Spanish, with a very clear pronunciation that we felt would be most comprehensible to students of Spanish, and easily understood by native-speakers. However, both native and non-native speakers felt this was not the case, so we re-recorded the test, this time using a young man who had grown up in Venezuela, but whose parents were Cuban. The Carribean influence was noticeable in his pronunciation. Before making this recording, we took the opportunity to make a few final revisions of items, and the latest version of the test was ready.

The Spanish IGT has not been a success in the same way as the English test. It does not clearly separate native and non-native speakers. Native speakers in the classes tested range from the high 90's to zero. It appears that a different kind of competence is being tapped. Whereas the typical items of the English test are relatively easy to do, requiring only a feeling for structural patterns of the language, most of the items on the Spanish test depend to some extent on the meanings of the sentences. In fact, in our attempts to strengthen some items that lacked discrimination, we can see in almost all cases that we added semantic clues. As a result it seems likely that the revised test measures general verbal ability instead of sensitivity to the grammar.

Some native speakers of Spanish do very well indeed; the highest score earned is 96. Non-native Spanish speakers also range widely, though usually not as far up or down as the natives. The best non-native score we have recorded is 90. These two scores are from a small group of language professionals who took the Spanish test. Their mean scores were:

	Natives	*Non-natives*
Mean	82.3	62.1

Comparable data for a similar group in the English test is:

	Natives	*Non-natives*
Mean	92.0	67.2

Data that are more revealing involve not postgraduate professionals with considerable linguistic sophistication and experience, but ordinary students. One group of students enrolled in the Santa Monica High School was especially attractive for research purposes: They were studying in a bilingual class where both Spanish and English are extensively used and in addition the classes are roughly balanced with comparable numbers of native speakers of each language. All students, native in English or Spanish, took the Integrative Grammar Tests in both English and Spanish. To cancel any practice effect giving an advantage to the language whose test was administered second, half the students took the English test first and half the Spanish. The results of this administration can be seen in Table 1.

The "N Adjusted" figure is reduced from N by the number of refusals, those who turned back a blank form. The means, ranges, and scores show that

TABLE 1

Performance Means of Santa Monica High School Students in Bilingual Social Studies Classes

Class	Native Speakers	Language & Sequence of Test	N	N Adjusted	Range	SD	Mean
1	English	English First	9	9	87.1	85-95	4.1
	Spanish	English First	15	7	4.4	0-11	4.3
	English	Spanish Second	9	8	23.8	12-37	9.2
	Spanish	Spanish Second	15	14	41.9	1-85	31.4
2	English	English Second	16	16	83.2	72-89	4.6
	Spanish	English Second	8	8	16.8	1-44	17.4
	English	Spanish First	15	15	21.7	8-44	10.2
	Spanish	Spanish First	10	10	44.2	1-83	27.0

Summary

	Combined English Means	Combined Spanish Means	First Test	Second Test
English Native Speakers	85.1	22.7	54.4	53.5
Spanish Native Speakers	10.6	43.1	24.3	62.6

both native-speaker groups do substantially better on tests in their own language, which is of course not surprising. But the English natives are in the 80's while the Spanish natives are in the 40's and 50's. The English natives are also higher in their second language than are the Spanish natives. The Spanish natives characteristically have wider ranges and standard deviation figures, showing greater diversity in performance. The only case where Spanish natives have a low SD is where the range is from 0-11, which allows a very limited space within which deviation is possible.

The Spanish natives do reasonably better on their second test, regardless of which language the test is in. This seems to be evidence that the Spanish test is measuring a general ability, in which practice is influential. The English natives actually have a small decline in their second test, showing a diminished interest in a test that quickly becomes routine. The test is so easy (mean of 87.1) that they tend to become inattentive, shown by a mean score that drops to 83.2 when the test comes second.

The correlations between test scores are very low and in no way meaningful. As Table 2 shows, getting a good score on one test does not mean a good score on the other. This suggests that first and second language skills are separate types of performance.

TABLE 2

Combined Correlation of Spanish and English ICT Scores

		Mean	SD	Range	r
English Native Speakers	English Test	84.2	6.0	66-90	.106
	Spanish Test	21.6	10.3	3-44	
Spanish Native Speakers	English Test	14.0	21.5	0-44	−.040
	Spanish Test	53.4	25.2	1-83	

The test just reported is a replication of a very similar test given a year earlier, using a preliminary version of the Spanish IGT. We won't report this test in detail other than to say that in general the relationships between English and Spanish tests are shown to be quite stable.

We do wish to cite a few correlation figures and the conclusions that seem to derive from the comparisons. Our observations are as follows:

Native English Speakers	*Native Spanish Speakers*
1. English test with course grades: r = .144. (Native language skills tested are unrelated to course achievement: test is focussed on applications of grammar, which everyone has mastered.)	1. Spanish test with course grades: r = .635. (Native language skills tested to some extent measure the same achievement as course grade test is apparently focussed on a more general verbal ability.)
2. Spanish test with course grades: r = .113. (achievement in a second language is likewise unrelated to course grades. Test seems *not* to measure whatever is being measured for class evaluation.)	2. English test with course grades: r = .405. (a higher, but nonsignificant correlation figure. Perhaps at least some students are not interested in the tests. There were 8 "refusals"—with zero scores, and an exhorbitant SD of 28.6!)
3. English test with Spanish test: r = .453. (a higher correlation, but still not very strong.)	3. Spanish test with English test: r = .455. (Closely similar.)

Another group of students took the Spanish IGT: students whose native language is Spanish enrolled in night classes in English language at the Evans Community School in downtown Los Angeles. This group provided an opportunity to correlate test performance scores with various measures of opportunity to master English. The best correlation was with years in school, with a coefficient of .639. This had not shown up at Santa Monica High School, because nearly all students had attended school the same number of years. The Evans data shows that IGT did not correlate positively with age, or with time spent in the U.S. These comparisons, made for two classes at Evans, showed a negative correlation (−.281 and −.592) for age—apparently older students were less successful—and (−.463 and −.139) for months spent in the U.S.A. It seems that the longer they stay in the U.S. (but still enroll for English instruction) the more likely they will have lower scores in their native language. Data for the performance in Spanish are as follows:

Class	*N*	*Mean*	*Range*	*SD*
1	9	68.1	31-93	18.5
2	17	71.4	7-92	21.9

The interesting details are the quite respectable means and the lack of homogeneity shown by the large ranges and standard deviations. These adult students outscore the high school students at Santa Monica High.

A more extensive administration of the Spanish IGT suggests other kinds of comparisons. One is with the Spanish version of the MLA Cooperative College Placement Test. The subjects were 153 first-term (new) student applicants to the Department of Spanish at UCLA seeking admission to classes the department offers. Actually it is a placement exercise, and the results of the examination are advisory only. This allows the examination to be given in a rather informal atmosphere, with a minimum of the tension usually associated with examinations. The Spanish IGT was added as a research caboose, given between the two administrations of the Spanish exam (LA and MA). Actually the IGT took about the 20 minutes that were needed to score the LA. On the basis of those scores, students were invited to take the more advanced MA. The break points are shown in Table 3.

TABLE 3
Testing Standards for UCLA Applicants

	LA	MA
Span 1	156 or lower	
Span 2	157 to 166	
Span 3	167 to 173	
Span 4		177 or lower
Span 5		178 to 183
Span 25		184 or higher

The Spanish IGT shows internal consistency by the data it produced in testing different language categories of examinees. The data, shown in Table 4, is just what we would predict from an acquaintance with the language background of the applicants. The interesting category is *Spanish surname*, claiming English as their native language and no effective knowledge of Spanish. In fact they average 50 per cent higher than the Anglo applicants as a group.

TABLE 4
Performance on Spanish IGT by Native-Language Categories

	N	Mean	Range	SD	SM
Native Spanish Speakers	10	66.8	25-97	18.95	5.99
Bilingual Spanish-English	6	62.2	44-34	16.99	6.93
Spanish Surname (Nat. Eng.)	15	32.5	2-63	18.57	4.90
Anglo (Scoring above 173)	43	33.4	16-82	14.48	2.21
All Anglo Examinees	116	23.4	1-82	12.70	1.18

At the time this is written, the applicants have been assigned to classes. (That is, 95 of them have; the other 58, for one reason or another, did not appear on the official grade sheets.) They have been distributed by the LA and MA test scores into classes at six levels: the first two years of Spanish instruction given at UCLA.

The comparison and correlation of the scores is most interesting. Summaries of these scores are shown in Table 5.

TABLE 5

Tables of Performance Means by Class

	M	LA (108)	MA (190)	IGT (100)	GPA (4.33)
Spanish 1	8	156.6	——	10.9	3.04
Spanish 2	17	160.5	180.3	18.1	3.41
Spanish 3	31	169.2	176.8	24.1	3.26
Spanish 4	14	174.4	176.3	23.5	2.83
Spanish 5	12	178.3	182.7	36.6	3.42
Spanish 25	18	181.6	187.7	52.5	2.91

The LA means are naturally consistent, since they are the basis of course assignment, directly up to Spanish 3 and indirectly for second-year classes. The MA means for Spanish 4, 5, and 25 are consistent, again because they were used in counseling placement for those classes. The first year means for MA scores do not agree with course placement, which is unexpected, since LA and MA do not disagree with each other by much? The Spanish IGT scores are consistent except for scores at course levels 3 and 4. Other means are in a logical relationship, and have quite large gaps between levels. The GPA means are not at all related to level, but there's no reason to expect them to be, since they are measures taken within each classroom and have no necessary relationship with other classes at the same level or classes on other levels. In fact, a consistent difference might be hard to explain. It looks like Spanish 4 and 25 are perhaps more strictly evaluated than are levels 2 and 5.

Table 6 shows the relative consistency of correlations (LA, MA, IGT, and GPA).

TABLE 6

Correlations Among Evaluation Measures

	Spanish 1			Spanish 2			Spanish 3		
	LA	MA	IGT	LA	MA	IGT	LA	MA	IGT
MA	——			.186			.790		
IGT	.264	——		.607	−.055		.628	.622	
GPA	.686	——	−.205	.113	−.721	.236	.224	.377	.038
	Spanish 4			Spanish 5			Spanish 25		
MA	.407			.440			.874		
IGT	.199	.684		.501	.351		.566	.292	
GPA	−.230	−.236	.046	.500	.010	.154	.366	.046	.224
				Combined Sections					
MA				.640					
IGT				.676	.023				
GPA				.091	−.203	.035			

The correlations at individual course levels run wild, ranging from LA-GPA .686 in Spanish 1 to a −.230 in Spanish 4. Figures from the combined sections show that none of the three tests correlates well with GPA; in fact the highest GPA correlation is a negative one (−.721 in Spanish 2).

The combined sections show that IGT correlates best with LA (.676) and not badly with MA (.623) which is not quite as good as LA with MA (.640). None of these predict GPA, which is a correlation we had hoped to find.

We set out on this project hoping that we might find a means of testing grammatical competence that could be applied across language boundaries. We have failed to do this, as least so far, which is a disappointment because the IGT idea looked hopeful. So, how do we account for the wide range of native-speaker performance: from 96% to 0?

In some instances, the Caribbean influence on our speaker's pronunciation may have played a part. In this dialect, the final /s/ tends to be aspirated, or pronounced more like a breathy /h/. Thus, *las*, for example, would sound more like *lah*. For examinees unfamiliar with this dialect, such a pronunciation could cause confusion on certain items, and a consequently lower score. Here's an example:

La sopera está en la mesa.
Las opera mañana por la mañana.

Here, the placement of the *s*—initial on *sopera* or final on *las*, is crucial to the identification of the second word of the sentence.

We think there is an even more important consideration in explaining the great variation in native-speaker scores. As I mentioned earlier, the features which differentiate test items are syntactic, phonetic, or semantic. The interpretation of the Spanish test items is heavily weighted toward the semantic component. The comprehension of only 36% of the items depends on a knowledge of Spanish grammar; interpretation of 16% of the sentences hinges on pronunciation differences. However, the comprehension of 48% of the items is based on semantics. This is not the case with the English test, 86% of whose items are based on syntax. Interestingly, one bilingual examinee commented that he somehow did not feel that the two tests measured quite the same thing. His intuition may be a comment on this dissimilarity between the two tests.

A knowledge of the grammar is much more intuitive and universal among native-speakers than is a knowledge of the vocabulary. An instrument that tests vocabulary assesses, to some extent, the sophistication of the examinees. In keeping with this theory, we have noticed that the scores of the native-speaker examinees seem to correlate with their level of education.

Revising the test so that more of the items are based on syntax might remedy this problem. For instance, items like these would depend more on a knowledge of Spanish grammar for their comprehension:

Ha sido muy lento el correo
Has ido muy lento al correo.

We think that by making the Spanish test more like the English version in what it measures, the two, used together, would provide a more accurate picture of bilingual balance.

An Alternative to the Cloze Test

Karen A. Mullen
University of Louisville

An editing task has been suggested as an indirect means by which to assess second language proficiency (Bowen 1978). The present study was conducted to see how well performance on this test compared to that on a writing task, an oral interview, and a cloze passage. The first two represent direct, integrative tests of proficiency, both of which require excessive time to administer and/or correct. The third is an indirect, integrative test, also requiring a relatively large amount of correction time. Data for this study were assembled in the form of two scores for the editing test (misidentification and non-identification), two scores for the cloze passage (exact-word and acceptable-word), and composition and interview evaluation scores. The results of the study show, first, the non-identification measure to be superior to the misidentification measure or a simple addition of the two in predicting performance on the other tests. Second, this measure correlates more highly with the direct tests than the cloze-acceptable measure does. Third, the score variance is larger, more item-facility and item-discrimination indices fall within the acceptable range, and the inter-item reliability is significantly higher for this editing measure than for the cloze-acceptable score. In view of the amount of time required to arrive at an acceptable-word cloze score and the less impressive performance of the cloze in relation to the editing measure, the editing task is recommended as a more satisfactory indirect, integrative test of second-language proficiency.

The cloze test is considered a relatively good instrument for testing second language proficiency; it has several features to recommend it, the most important being ease of construction (Pack, 1973) and high correlations with direct tests of second language proficiency (Hinofotis 1977, Mullen 1979c). It also has been shown that differences among subjects taking the test rather than differences between the two commonly accepted means of correction or differences due to passage difficulty make the most significant contribution to explaining the score variance (Mullen 1979c). However, it has been suggested that correction be done by the acceptable-word method since it tends to produce scores which not only have a larger variance but also correlate more highly with criterion tests (Hinofotis 1977, Oller 1972, Mullen 1979c). Several procedures for determining acceptability have been proposed, and studies have shown that the resulting scores do not differ significantly in the degree to which they correlate with criterion tests (Oller 1972). But one feature common to all acceptable-word procedures is that they must be rigorously defined and applied and consequently much time is required for the correction. A second worrisome feature of the cloze is the attitude subjects have toward it, considering it an inappropriate, even an unfair test, and therefore do not accept it as a *bona fide* means by which to evaluate their proficiency. Because of these practical problems with the cloze—that of scoring time and face validity—the editing test (Bowen 1978) might be a suitable alternative. Like the cloze, it requires that the subject process the syntax and semantics of the text in order to restore it to its original form; in the cloze, this is done by supplying what is missing, in the editing, by deleting what is alien.

---

* This is a slightly expanded and revised version of a paper presented at the 1979 TESOL Convention in Boston, Massachusetts.

In view of the promising correlations to a criterion test such as the Michigan Battery in Bowen's study and the ease of scoring, a study was conducted at the University of Louisville to see whether performance on an editing test compared favorably with performance on a cloze test and on two direct tests of second language proficiency—an oral interview and a writing task.

1. Method

1.1 Subjects

Of the fifty-four foreign students seeking admission to the University of Louisville who served as subjects, some had finished intensive English programs and were applying for admission to a degree program while others had studied at a local community college and were hoping to transfer to the University. A few were foreign students, recently graduated from high school, who were seeking admission to an undergraduate program.

1.2 Materials and Design

Four tests were administered to the subjects: an editing test, a cloze test, a writing task, and an oral interview. Some subjects had recently taken the TOEFL prior to the testing and their scores were included in the study.

The editing test was based on a seventh-grade reading level passage taken from the Science Research Associates Reading Laboratory IV. The randomly inserted words were assembled by taking the first word in the second line of every left-handed page of an ESL text of comparable reading difficulty. They were inserted one by one at random points in the text, determined by the throw of three dice. To keep the passage length closely similar to that of the cloze test and to make the average distance between insertions as close as possible to the distance between blanks in the cloze test, the positions available for insertion ranged from after the fifth to after the fifteenth word in the base text. Subjects were told to cross out the fifty words which did not seem to fit in the text. A sample test is illustrated in Figure 1. It was corrected by first overlaying a stencil to reveal all words except those which were randomly inserted; any crossed-out words among these were counted as misidentification errors. A second stencil disclosing the randomly inserted words was then placed over the passage; any word which a subject failed to cross out was counted as a non-identification error. This follows the procedure described by Bowen.

Figure 1
A Sample Editing Test

DIRECTIONS

This is a test designed to measure your proficiency in English. As you read the passage below, you will find that some words do not seem to belong in the text. Cross out the words which you think do not fit in the sentence or the passage. You should cross out fifty words.

Example: Since man first appeared on earth he ~~allowed~~ has had to solve certain problems of survival, such ~~early~~ as hunger, thirst, and cold.

In the long history of man's inventiveness pleased, discoverers seem to fall into salary two classes. The first is the ingenious person who sets out interesting to find the solution to a problem. The second is choice the lucky one who appears to would stumble upon something by accident. But we should be clear what we mean never by accident, for the accidental aspect of many great discoveries and is that something unusual has happened when there is an observant person present who notices took what has happened and sets he to work to find out why.

The cloze test was also based on a seventh-grade reading level passage. Every 10th word was deleted and replaced with a blank of uniform size. Fifty words were deleted. Subjects were told to supply the best word they could think of to fit the grammar and meaning of the text surrounding the blank. Responses matching the deletions from the original passage comprised the exact-word score. Responses considered as appropriate alternatives by two judges comprised the acceptable-word score. In cases where the two judges could not agree, the response was submitted to a third judge.

In order to control for fatigue and a possible practice effect, half of the group was given the editing test first, the other half the cloze test first. Although there was no time limit, no subject took longer than an hour on each test. Following the first test, subjects wrote a composition from a choice of topics; they were told that their compositions would be evaluated for structure, organization, quantity, and vocabulary. To arrive at an evaluative measure, a Likert-type scale divided into nine equal segments was provided for each of the four criteria mentioned above. The nine segments were labelled *poor*, *fair*, *good*, *above average*, and *excellent* or in between. Guidelines for determining designations were available for reference (Mullen 1977a). Four readers evaluated each composition on each of the four scales. The designations were converted to numerical values ranging from 1 for *poor* to 9 for *excellent*. The final score was computed as a percentage based on the total accumulated points in relation to the maximum attainable.

Forty-three of the fifty-four subjects were also interviewed. The procedure for conducting the interview and for evaluating the performance followed that in other studies (Mullen 1977b, 1978). The scales were of comprehension, fluency, control over structure, and pronunciation. Guidelines were available for determining position along the scale. Values were assigned in the same manner as for the composition, and a percentage score was calculated, again based on the number of points accumulated in relation to the maximum attainable. Four persons evaluated the interview performance.

2. Results and Discussion

Relevant descriptive measures are reported in Table 1. The mean cloze scores and the mean composition and interview scores are not significantly different from those derived in a larger study of 154 subjects using the same cloze test, composition task, and interview procedure. Mean TOEFL scores available for only a small segment of the subject group are also reported as a point of reference.

TABLE 1

Mean Score and Standard Deviations for Editing Test, Cloze Tests, TOEFL, and Composition and Interview Evaluations.

Tests	Mean	S.D.	N
Editing—Non-identifications (# wrong)	19.44	13.62	54
Editing—Misidentifications (# wrong)	17.98	14.20	54
Cloze—Acceptable (# right)	21.50	10.45	54
Cloze—Exact (# right)	14.90	8.27	54
TOEFL—Listening Comprehension	47.27	5.56	18
TOEFL—Structure and writing Ability	46.[illegible]	7.35	14
TOEFL—Reading and Vocabulary	48.21	7.06	14
TOEFL—Total	468.40	51.66	22
Composition—% of possible points	54.50	21.98	54
Interview—% of possible points	62.41	22.30	43

In considering how well performance on the editing test correlates with performance on the cloze test and the two direct proficiency tests, it is apparent that the non-identification measure produces a higher correlation than the misidentification measure (Table 2). The effect of adding the latter to the former to compute an overall score generally results in depressed correlations. Clearly, the non-identification score is a better measure. It is also important to note that

TABLE 2

Pearson Product Moment Correlations of Scores of the Editing Test with the Cloze Tests, the TOEFL, and the Composition and Interview Evaluations.

Test	Misident.	Non-ident.	Misident. + Non-ident.	N
Cloze Exact	−.25*	−.73***	−.5[illegible]***	54
Cloze Acceptable	−.39**	−.85***	−.74***	54
TOEFL—Listening	−.32	−.60**	−.53*	18
TOEFL—Str. & Wtg.	−.55**	−.60**	−.64**	14
TOEFL—Rdg. and Voc.	−.05	.16	.07	14
TOEFL—Total	−.25	−.43*	−.43*	22
Composition	−.39**	−.82***	−.71***	54
Interview	−.15	−.74***	−.53***	43

* $p < .05$ ** $p < .01$ *** $p < .0001$

the non-identification score correlates more highly with the acceptable-word method of cloze-scoring (−.84) than with the exact-word score (−.73). In comparing the acceptable-word measure with the non-identification measure, the latter correlates more highly with the composition and the interview evaluations (Table 3).

There are some other important features of the non-identification measure when compared to the acceptable-word measure, reflected in Tables 4 and 5. One is that the score variance is larger, indicating that the editing test can spread scores further out along the scale, thereby allowing for finer discrimination in levels of proficiency. An analysis of responses for each item in the two tests reveals that for the editing test, the mean item facility is within a more acceptable range (61%), that no item falls below a facility index of 30% or above 85%, and that fewer items fall below the recommended discrimination

TABLE 3

Pearson Product Moment Correlations of the Non-Identification Scores of the Editing Test and the Acceptable-Word Scores of the Cloze Test with the Composition and Interview Evaluations.

Test	Cloze Acceptable	Editing # Identified	N
Composition	.77	.82	54
Interview	.67	.74	43

TABLE 4

Item Facilities (IF) and Item Discriminations (ID) Based on Corrected Item-Total Correlations for the Cloze Acceptable Scoring and the Scoring of the Correctly Identified Insertions on the Editing Test.

Item #	Cloze IF	Cloze ID	Editing IF	Editing ID	Item #	Cloze IF	Cloze ID	Editing IF	Editing ID
1	.83	.14	.69	.61	26	.33	.52	.54	.36
2	.80	.53	.80	.45	27	.74	.07	.70	.67
3	.48	.60	.37	.46	28	.52	.49	.44	.43
4	.46	.51	.70	.53	29	.46	.56	.83	.68
5	.44	.59	.70	.47	30	.63	.53	.43	.62
6	.22	.26	.78	.39	31	.37	.54	.56	.64
7	.33	.38	.44	.61	32	.70	.59	.54	.44
8	.35	.55	.76	.37	33	.35	.19	.81	.54
9	.59	.48	.76	.59	34	.28	.56	.81	.68
10	.44	.48	.81	.62	35	.01	.27	.54	.60
11	.11	.12	.52	.70	36	.50	.64	.30	.29
12	.70	.41	.52	.41	37	.35	.45	.65	.54
13	.31	.44	.83	.49	38	.15	.37	.78	.65
14	.83	.42	.72	.67	39	.11	.34	.31	.50
15	.38	.50	.65	.38	40	.48	.10	.35	.34
16	.37	.26	.61	.74	41	.72	.48	.67	.38
17	.69	.71	.54	.58	42	.41	.48	.39	.70
18	.11	.48	.50	.72	43	.15	.35	.80	.62
19	.93	.45	.44	.42	44	.33	.42	.69	.68
20	.37	.32	.72	.51	45	.20	.38	.80	.67
21	.07	.48	.76	.68	46	.65	.54	.43	.49
22	.39	.45	.48	.51	47	.44	.53	.57	.67
23	.61	.53	.54	.61	48	.20	.33	.70	.66
24	.54	.51	.48	.53	49	.13	.38	.50	.47
25	.55	.31	.70	.73	50	.30	.60	.52	.62

TABLE 5

Comparison of Mean Scores, Item Facilities (IF), Item Discriminations (ID) Based on Item-Total Correlations, Related Standard Deviations, and Inter-Item Reliabilities for the Cloze Acceptable Measure and the Editing Test Measure Based on Number of Insertions Correctly Identified.

Measure	Cloze Acceptable	Editing (# Identified)
Mean Score	21.50	30.48
S.D. of Mean Score	10.45	13.62
Mean IF	.43	.61
S.D. of IF	.22	.15
Number of Items with IF below .30	12.00	0
Number of Items with IF above .85	1.08	0
Mean ID	.43	.56
S.D. of ID	.14	.11
Number of Items with ID below .30	8.00	1.00
Inter-Item Reliability	.92	.96

index of .30. A third feature is that the coefficient of inter-item reliability is higher for the editing test than for the cloze test. This difference is significant at the .03 level.

In summary, the results of this study show the non-identification score to be a better measure than either the misidentification score or a simple addition of the two. As measured by correlations with the composition and the interview, this measure is superior to the acceptable-word cloze measure. Moreover, the data show that the measure is superior in terms of score variance, item facility, item discrimination, and inter-item reliability. Finally, the test is easy to construct and scoring requires less time.

Scoring Procedures For Use On Cloze Tests

J. Charles Alderson

University of Michigan

Recent years have seen a growth in interest in the use of the cloze procedure in EFL proficiency testing. Some research seems to have shown that the cloze is a valid measure of such proficiency and that, as a so-called integrative test, it relates closely to other supposedly integrative measures of comprehension and production. Doubts have centered on the scoring procedure to be used on the cloze test, but previous research has produced somewhat ambiguous results. The present tendency is to use the exact word procedure.

This paper will present the results of an investigation into the use of five different scoring procedures, which will show that for the purpose of EFL proficiency testing, the semantically acceptable procedure is the most appropriate. Evidence will be presented which shows that this procedure drastically reduces the effect of confounding variables like deletion frequency and text difficulty, and that it results in better discrimination of native and non-native speakers of English. The implications of this finding for an analysis of what the cloze test measures will be discussed, and it will be suggested that a radical revision of the uses of the procedure is necessary.

Research into the use of different scoring procedures on cloze tests with native speakers has generally come to the conclusion that the exact word procedure (where only replacements of the exact word deleted are counted as correct) is the most valid and practical way to score cloze tests. Taylor (1953), Rankin (1957) and Bormuth (1965) all find high correlations between the exact method, and methods allowing "any synonym" or "good enough answers", although the means of the latter scores were slightly higher. Bormuth (1965) found that the nearer the response came to the exact word, the higher were the correlations with other measures of text comprehension. However, using a "grammatically correct score", Hafner (1964) found a correlation of only .61 with the exact word score. Fillenbaum et al. (1963) suggest that a form class score—giving credit for replacements from the same form class as the deletion—might measure sensitivity to the close grammatical environment of the deletion, whilst the exact word score might depend more on sensitivity to remote semantic features of the discourse. In general, however, the practice with native speakers of English has been to use the Exact Word scoring procedure.

With non-native speakers there has been a greater tendency to use scoring procedures other than the exact word; in fact, it is impossible to compare many of the research results using the cloze procedure because the researchers have used different scoring procedures. In general, there is a feeling among EFL and ESL researchers and testers that the exact word method is too harsh on a non-native speaker of the language. As Oller (1972b) points out, given the sentence, *The ______ climbed up the tree*, a response of *horse* or *elephant* instead of *monkey* is of a different order, i.e. displays a different level of language competence, from responses like *table* or *with*. Scoring procedures that have been used with non-native speakers have included the communality of response score (Carroll et al. 1959) and the clozentrophy score (Darnell 1968). In the former case, it was found that there was a high intercorrelation with the exact word score, and lower correlations with validating criteria than the exact word score achieved, and so the procedure was rejected. In the latter case however, a high validating correlation was achieved; although no intercorrelation with the exact procedure was possible, the clozentrophy was held to be a promising alternative to the exact word procedure.

Anderson (1972) got a high (.99) correlation of the exact word procedure with a synonym scoring procedure, but unfortunately the latter procedure was in any case weighted in favour of the exact word. The classic studies of the use of different scoring procedures with non-native speakers are by Oller (1972b); Oller, Atai and Irvine (1974); and Stubbs and Tucker (1974). Oller (1972) found that he got a higher validity coefficient for his acceptable word scoring procedure than for the exact word procedure (.83 and .75 respectively), and concluded that the acceptable word procedure was the best. Unfortunately he did not report intercorrelations of the procedures. Oller, Atai and Irvine (1974) found an intercorrelation of .94 between the exact word and the acceptable word procedure, and claimed that all validating correlations with the TOEFL were identical for both procedures, and therefore recommended use of the exact word procedure. However, they got a correlation of .75 with the acceptable procedure and dictation, against .69 using the exact word procedure and so one might have expected them to reach the opposite conclusion, given the assumption that cloze is more valid when it correlates higher with supposedly integrative tests like dictation. Stubbs and Tucker (1974) found an intercorrelation of .97 between the exact word method and an acceptable word procedure, and recommended the exact word method, ignoring the fact that the acceptable word method attained higher validating correlations than the exact word. In fact, the differences between the correlations of the two procedures with the criterion entrance test of the American University of Beirut was of the same order as that found by Oller (1972b) with his criterion, which led him to the opposite conclusion, i.e., use the acceptable word method.

The studies reported are open to criticism on several grounds. In some cases the scoring procedures used are weighted in favour of the exact word, thus biasing the results. In other cases, only partial data are reported. Often the researchers merely report correlation coefficients, and not even the means and standard deviations of their tests. Conclusions are based on partial evidence —either on the intercorrelations of the procedure under consideration or on the validating correlations, but rarely on both. In some cases, as mentioned above, clear counterevidence is ignored.

In addition, and perhaps more seriously, there is a conspicuous failure to account for the similarity of the correlations that are achieved. Why does the exact word procedure result in apparently the same validity coefficients as the semantically acceptable word procedure? Do the similarity of these coefficients and the high intercorrelation of procedures mean that in essence both procedures are measuring the same thing? At first sight this would seem to be counterintuitive, since the ability to restore the exact word may well depend, for certain items at least, on the ability to detect distant relationships amongst elements in text, whereas the ability to restore with an acceptable word is at least arguably less dependent on discourse sensitivity. Intuitively, at least, one would expect different scoring procedures to result in different measures of some aspect of linguistic proficiency—as suggested by, for example, Fillenbaum et al. (1963) and Bormuth's results (1965), both reported above, as well as Hafner's (1964) finding that a grammatically correct score was less related to reading achievement and intelligence than the exact word score.

As part of an investigation into the effect of several variables on cloze test performance, certain procedures used to score cloze tests were examined. Specifically, five scoring procedures were developed, as described below:

1. The exact word procedure
2. The semantically acceptable procedure (SEMAC)
3. The same form class procedure (IDFC)
4. The acceptable form class, same grammatical function procedure (ACFC)
5. The grammatically correct procedure (GRCO)

1. *The exact word procedure.* In this procedure, minor misspellings are not counted as incorrect.

2. *The semantically acceptable procedure (SEMAC).* The problem with this procedure is the decision as to whether a response is acceptable or not. Is it, for example, acceptable to replace the name *Mr. Vaughan* with the name *Mr. Smith,* although no Mr. Smith has been mentioned in the text and will not be mentioned later? One marker may consider stylistic infelicities to be unacceptable, whilst another may find them acceptable. Rather than develop elaborate instructions for acceptability scoring (see, for example, Clarke and Burdell, 1977), which it was felt would never be applied in practical cloze testing, it was decided to compare markers using the any acceptable word procedure, to see if practical agreement could be reached without any analytical scheme. The nature of this investigation is reported elsewhere but, essentially, it was found that ten native speaker markers agreed with each other on acceptability, with correlations ranging from .93 to .98, and that they agreed with themselves after one month's interval, again between .93 and .98. Seven non-native speaker markers achieved the same levels of agreement and reliability, and agreed with the native speaker markers at between .84 and .93. Thus, contrary to frequent claims and suppositions, it is possible to achieve high agreement on acceptability. This author's marking for acceptability, computer-scored to achieve 100%

objectivity, agreed with the native speakers' combined judgments at .99, and at .96 with non-native speakers. The procedure was such as to disallow *Mr. Smith* for *Mr. Vaughan* in the case mentioned above, nor was a *she* admitted for a *he* where no female person had been mentioned; however, allowable replacements for *yellow* in the sentence *A* ______ *sun wallowed high above the mountains* included words like *blazing bright hot large lazy lovely magnificent nice red setting* and *watery*.

In an attempt to develop a scoring procedure measuring grammatical sensitivity, the following procedures were evolved:

3. *The same form class procedure* (IDFC): If a response is a member of the same form class as the deleted word it is correct; if not it is incorrect—the semantic fit of the response is ignored. Problems of form class classification arose—is *as* the same form class as *like* or *than*? Are *he was due to go* and *he was able to go* equivalent? A more serious problem was the severity of error. In the context *He gazed dreamily at the Baptist preacher* and *they fought in order to sit next to Monroe* the replacement of *Baptist* by *yellow* (different form class) is of a different order from the replacement of *next* by *yellow* or, indeed, by *beside*, presumably the same form class. For these reasons, the following modified scoring procedure was also used.

4. *Acceptable form class, same grammatical function* (ACFC): If a response was from a form class which was acceptable in the context of the item, it was scored as correct, provided that the response had the same grammatical *function* as the deleted word. Grammaticality of concord, number, tense, etc. was ignored. Thus, unlike the previous procedure (IDFC), *The shop window was broken* was an acceptable replacement for *The big window was broken* since *shop* and *big*—different form classes—both fulfill the function of modifier. Examples of this procedure, where one alternative would not be regarded as being an acceptable replacement for the other, are as follows:

1. The $\begin{matrix}\text{very}\\ \text{nice}\end{matrix}$ old gentleman was walking down the street.
2. The man $\begin{matrix}\text{and}\\ \text{with}\end{matrix}$ his dog can be seen in the distance.
3. He came $\begin{matrix}\text{when}\\ \text{and}\end{matrix}$ she went.
4. The $\begin{matrix}\text{farmers}\\ \text{they}\end{matrix}$ sold all their food.

The problem with this procedure is that it penalises grammatically correct responses (as 1–3 above) and allows grammatically incorrect responses *A birds was singing happily*. In order to get at a more complete grammatical sensitivity, this procedure was supplemented by the final procedure.

5. *The grammatically correct procedure* (GRCO): Any response which fits the syntax of the context is correct. It must agree in number, concord, etc. with the environment and be from an acceptable form class, but it need not have the same grammatical function as the deletion. As far as possible, semantic relatedness and appropriacy are ignored.

1. The study

These five scoring procedures were applied by computer to the responses to 12 different cloze tests produced by deleting every 6th, 8th, 10th and 12th word from three texts of varying difficulty (easy, medium and difficult respectively). (The difficulty of the texts was determined by a panel of EFL teachers and several readability formulae (Smog, Fog, Flesch and Dale-Chall.) The tests were given to 360 native speakers of English and 360 non-native speakers studying in the UK at tertiary level, which, since each subject took only one cloze test, gave 30 cloze tests per group. In addition the non-native speakers were given the English Language Battery (ELBA) (Ingram, 1964, 1973) which contains seven subtests, and two dictation tests, one easy, one difficult.

2. Results

This paper will be primarily concerned with the non-native speaker results. In point of fact, however, virtually all results were the same for native and non-native speakers.

As can be seen from Table 1, the three texts were ranked in the same order, regardless of scoring procedure.

TABLE 1

Mean scores for cloze tests, non-native speakers

Deletion rate 6					
	Exact	Semac	Grco	Idfc	Acfc
Easy	30.3(1)	42.4(1)	45.9(1)	44.3(1)	45.7(1)
Medium	20.1(2)	30.3(2)	39.1(2)	38.9(2)	39.8(2)
Difficult	14.8(3)	24.9(3)	37.8(3)	31.2(3)	33.0(3)
Deletion rate 8					
	Exact	Semac	Grco	Idfc	Acfc
Easy	29.8(1)	40.7(1)	44.6(1)	42.7(1)	44.7(1)
Medium	19.9(2)	31.3(2)	40.4(2)	37.5(2)	41.2(2)
Difficult	10.0(3)	21.6(3)	31.3(3)	[illegible](3)	27.8(3)
Deletion rate 10					
	Exact	Semac	Grco	Idfc	Acfc
Easy	30.8(1)	40.3(1)	44.2(1)	44.6(1)	44.7(1)
Medium	23.4(2)	31.6(2)	38.8(2)	39.5(2)	41.2(2)
Difficult	9.4(3)	19.0(3)	30.7(3)	27.0(3)	28.7(3)
Deletion rate 12					
	Exact	Semac	Grco	Idfc	Acfc
Easy	28.7(1)	38.7(1)	44.4(1)	42.1(1)	44.5(1)
Medium	21.0(2)	28.5(2)	36.6(2)	37.8(2)	38.5(2)
Difficult	14.6(3)	24.7(3)	35.7(3)	30.1(3)	31.4(3)

It is more difficult for non-native speakers to respond with replacements fulfilling the same grammatical function as the deletion on a medium text than on an easy text, and easier to supply a grammatically correct replacement on a medium text than on a difficult text. No scoring procedure, however permissive, seems insensitive to the differences in text difficulty.

Although an inspection of the means in Table 1 suggests that different means were obtained on the same test by different scoring procedures, t-tests for differences between means were calculated to see if different scoring procedures resulted in *significantly* different means on any given test. On virtually all the comparisons, this proved to be the case. The only exceptions, i.e., where no significant differences were found, were on the easier texts, among grammatical scoring procedures (ACFC vs GRCO). Thus it can be concluded that using a different scoring procedure results in different mean scores on virtually all cloze tests, regardless of deletion frequency.

The order of difficulty was: Exact, SEMAC, IDFC, ACFC, GRCO; it is easier to provide a grammatically correct response than a response with the same grammatical function or from the same form class as the deletion. As the constraints on replacement increase, so does the difficulty of closure.

Table 2 shows the correlation coefficients for the intercorrelations of the scoring procedures.

Table 2 shows that although the scoring procedures resulted in different mean scores, they show a high degree of agreement with each other on the ranking of subjects—of 120 coefficients, the lowest is .71, whilst the highest is .99 and 47 out of 120 are at least .90. The closest relationship is among grammatical procedures, whilst the lowest relationship is between exact and grammatical procedures. The ability to predict grammatical function or to respond with a grammatically correct response would appear not to be closely related to the ability to identify the exact word deleted—the former, ACFC, being taken to be the lowest form of grammatical sensitivity whilst the latter could be considered the highest form of sensitivity to style, author's intention and so on, if these procedures really do measure different abilities. However, even these two extremes—ACFC and EXACT—relate to each other at about .80, suggesting that, on the whole, the procedures might be measuring much the same ability.

Although the exact word score correlates best with the semantically acceptable procedure (SEMAC), the SEMAC does not always correlate highest with the exact word procedure, (tests D08, E12, E06, E10, M08, M10) sometimes showing a closer relationship to the grammatical procedures. These findings are somewhat surprising in view of the claim that cloze tests are integrative tasks, and the parallel claim that the semantically acceptable procedure is the most valid way of scoring responses. It remains to be seen whether different procedures measure different abilities.

As has been reported elsewhere (Alderson, submitted) varying the deletion frequency on a cloze test results in significant differences in mean scores for some tests; the differences are not always in the expected direction (such that one would expect 6 to be harder than 8 to be harder than 10 to be harder than 12), and they vary from text to text. However, using any scoring procedure other than the exact word greatly reduced the effect of the deletion frequency variable; whereas the exact procedure resulted in significant differences on difficult and easy texts, when the SEMAC was used, no difference was found on

TABLE 2

Intercorrelation of scoring procedures (Spearman rho)

Difficult text		Semac	Greo	Idfc	Acfc
D06	Exact	.90	.71	.73	.75
	Semac		.71	.76	.77
	Greo			.87	.92
	Idfc				.96
D08	Exact	.92	.85	.89	.83
	Semac		.92	.95	.90
	Greo			.96	.97
	Idfc				.98
D10	Exact	.92	.74	.78	.74
	Semac		.85	.89	.86
	Greo			.97	.99
	Idfc				.98
D12	Exact	.95	.84	.84	.83
	Semac		.91	.90	.88
	Greo			.90	.89
	Idfc				.99
Medium text					
M06	Exact	.94	.84	.76	.77
	Semac		.92	.81	.81
	Greo			.89	.90
	Idfc				.99
M08	Exact	.87	.83	.74	.72
	Semac		.90	.81	.82
	Greo			.88	.87
	Idfc				.92
M10	Exact	.89	.84	.80	.86
	Semac		.95	.88	.89
	Greo			.93	.94
	Idfc				.95
M12	Exact	.95	.86	.89	.89
	Semac		.92	.90	.90
	Greo			.94	.94
	Idfc				.99
Easy text					
E06	Exact	.88	.84	.73	.72
	Semac		.92	.75	.80
	Greo			.82	.87
	Idfc				.97
E08	Exact	.94	.86	.88	.85
	Semac		.93	.86	.83
	Greo			.81	.84
	Idfc				.95
E10	Exact	.86	.81	.81	.81
	Semac		.94	.86	.87
	Greo			.87	.88
	Idfc				.99
E12	Exact	.85	.79	.88	.84
	Semac		.90	.89	.91
	Greo			.90	.88
	Idfc				.94

the easy or medium text, and only two of the six contrasts on the difficult text were significant.

As a further aspect of the investigation of the five procedures, item facility and discrimination indices and test reliabilities were computed and the procedures were compared. From Table 3 it can be seen that the SEMAC results in the highest number of acceptably difficult items on both medium and difficult texts.

TABLE 3
Number of items with a facility index between 20% and 80%
(Total number of items = 50)

	Exact	Semac	Greo	Idfc	Acfc
D06	24	41	29	35	32
D08	18	39	38	36	39
D10	15	35	37	36	35
D12	20	38	33	33	33
M06	26	37	28	27	25
M08	25	34	17	14	12
M10	26	31	22	19	16
M12	33	33	29	21	20
E06	33	17	4	8	6
E08	34	22	9	12	10
E10	22	19	8	7	7
E12	32	23	9	14	8

The grammatical scoring procedures produce more acceptable item difficulties than the exact procedure on the difficult text, but on the other two texts they are clearly inferior to either the SEMAC or the EXACT. In terms of item discrimination, however, the SEMAC always results in the highest number of discriminating items, regardless of text, and even the grammatical procedures give more discrimination than the exact word procedure on both difficult and medium texts (Table 4).

TABLE 4
Number of items with a Discrimination Index (E_{1-3}) of above +.2

	Exact	Semac	Greo	Idfc	Acfc
D06	18	30	30	28	26
D08	19	41	38	34	37
D10	13	35	34	34	35
D12	18	35	31	33	32
M06	24	33	29	26	25
M08	24	33	25	19	18
M10	21	28	23	16	19
M12	23	31	33	29	28
E06	19	20	10	13	9
E08	26	25	13	15	13
E10	20	23	18	11	12
E12	23	26	19	12	9

The SEMAC is clearly superior to other procedures, except on very easy texts, in terms of item efficiency. If the exact word procedure is being used, then the text should be an easy one; however, if the text is of medium or rela-

TABLE 5

Reliabilities of cloze tests (number of items = 50)

		Exact	Semac	Greo	Idfc	Acfc
D06	mean	14.8	24.9	37.8	31.2	33.0
	s.d.	5.3	7.2	7.3	6.6	6.8
	s.e. as %	6.5	5.3	3.5	3.8	3.8
	KR 20	.76	.82	.88	.80	.83
D08	mean	10.0	21.6	31.3	25.6	27.8
	s.d.	4.9	9.6	10.0	8.4	8.8
	s.e. as %	8.9	8.1	5.8	6.0	5.8
	KR 20	.81	.91	.92	.88	.89
D10	mean	9.4	19.0	30.7	27.0	28.7
	s.d.	3.7	8.1	9.7	9.0	9.1
	s.e. as %	7.3	7.8	5.8	6.0	5.8
	KR 20	.69	.87	.91	.90	.90
D12	mean	14.6	24.7	35.7	30.1	31.4
	s.d.	5.2	8.8	8.3	7.9	8.0
	s.e. as %	6.5	6.5	4.2	4.6	4.6
	KR 20	.80	.89	.89	.88	.88
M06	mean	20.1	30.3	39.1	38.9	39.8
	s.d.	5.8	8.2	6.9	5.9	5.8
	s.e. as %	5.3	5.0	3.2	2.8	2.7
	KR 20	.80	.88	.86	.81	.81
M08	mean	19.9	31.3	40.4	37.5	41.2
	s.d.	5.3	7.9	6.5	4.5	5.1
	s.e. as %	4.9	4.6	2.9	2.2	2.3
	KR 20	.76	.87	.85	.74	.81
M10	mean	23.4	31.6	38.8	39.5	41.2
	s.d.	4.8	6.4	5.8	4.8	4.6
	s.e. as %	3.8	3.7	2.7	2.2	2.0
	KR 20	.69	.80	.81	.74	.75
M12	mean	21.0	28.5	36.6	37.8	38.5
	s.d.	5.8	8.0	7.9	7.3	7.4
	s.e. as %	5.1	5.1	4.0	3.5	3.5
	KR 20	.78	.88	.89	.88	.89
E06	mean	30.3	42.4	45.9	44.3	45.7
	s.d.	4.3	5.8	4.3	4.1	3.9
	s.e. as %	2.6	2.5	1.7	1.7	1.6
	KR 20	.53	.84	.82	.77	.79
E08	mean	29.8	40.7	44.6	42.7	44.7
	s.d.	6.3	6.3	4.1	3.7	3.8
	s.e. as %	3.9	2.8	1.7	1.6	1.5
	KR 20	.79	.84	.75	.67	.71
E10	mean	30.8	40.3	44.2	44.6	44.7
	s.d.	5.6	7.0	5.0	4.1	4.1
	s.e. as %	3.3	3.1	2.1	1.7	1.7
	KR 20	.79	.89	.84	.77	.77
E12	mean	26.7	38.7	44.4	42.1	44.5
	s.d.	6.5	7.9	5.0	4.4	3.8
	s.e. as %	4.4	3.7	2.1	1.9	1.5
	KR 20	.80	.90	.85	.75	.75

tively high difficulty—usually the case in tests of EFL proficiency and reading comprehension—then the semantically acceptable procedure is the best procedure to use.

Table 5 shows means, standard deviations, standard errors of the means as percentages of the mean, and Kuder Richardson 20 reliabilities for all scoring procedures.

The dispersion achieved by the grammatical procedures is not as great as that achieved by the SEMAC or the Exact. The exact word procedure gives a distribution which is either greater than or equal to that of the SEMAC on all three texts, in relative terms (as a proportion of the mean) although in absolute terms the standard deviations of the exact word procedure remain remarkably constant at around 5.0, regardless of the variation in the mean over texts. In absolute terms the SEMAC almost always results in the highest standard deviation. However, the exact word procedure also results in a larger standard error than the SEMAC. Above all, the semantically acceptable procedure is always more reliable (as measured by KR 20) than the exact word procedure and at least as reliable as the grammatical procedures. In fact, the SEMAC's reliability coefficient is much more consistent than the exact word over deletion frequencies, ranging from .80 to .91 compared with the exact word procedure's range of .53 to .81. Here again, the SEMAC reduces the variation in cloze test performance caused by changes in the text and deletion frequency variables.

Correlation coefficients were calculated for the cloze tests scored by the five different procedures and the external measures of EFL proficiency—the easy and difficult dictation tests and the ELBA. The latter consists of seven subtests: Sound Recognition (1), Intonation (2), Sentence Stress (3), General Listening Comprehension (4), Grammar (5), Vocabulary (6) and Reading Comprehension (7).

Table 6 shows the correlations of each scoring procedure with the overall scores on ELBA, taken to be the best composite measure of EFL proficiency.

Consistently the SEMAC produces amongst the highest correlations—it almost always correlates higher than the exact word procedure, and with two

TABLE 6

Correlation of Scoring Procedures with ELBA Total

	Exact	Semac	Groo	Idfc	Acfc
Easy text					
E06	.59	.74	.60	.44	.45
E08	.70	.69	.61	.50	.46
E10	.65	.74	.75	.63	.65
E12	.67	.77	.72	.73	.71
Medium text					
M06	.86	.88	.8[illegible]	.67	.68
M08	.68	.77	.74	.51	.50
M10	.57	.74	.75	.70	.65
M12	.73	.78	.75	.70	.69
Difficult text					
D06	.51	.67	NS	.43	.43
D08	.82	.87	.73	.80	.74
D10	.79	.83	.79	.83	.82
D12	.77	.85	.68	.72	.70

exceptions it also correlates higher than the GRCO. In fact, the GRCO results in better correlations than the Exact procedure on deletion rates 10 and 12 on Easy and Medium texts. The SEMAC also gives the most consistently high correlation with EFL proficiency, ranging from .67 to .88 (Exact word procedure ranges from .51–.86), thereby reducing the effect of deletion frequency and text changes. Thus one way to stabilise coefficients is to use the semantically acceptable procedure rather than the exact word procedure. The grammatical scoring procedures, in particular the form class procedures ACFC and IDFC, do not relate very closely to the ELBA criterion. Table 7 below allows one to compare each scoring procedure with the ELBA subtests and dictation (differences between deletion frequencies are ignored).

TABLE 7

Correlations of Cloze Scores with ELBA and dictation by Text

	ELBA							Dictation	
	1	2	3	4	5	6	7	E	D
Difficult text									
Exact	.36	.49	.25	.55	.71	.57	.54	.50	.54
Semac	.50	.57	.30	.67	.79	.73	.63	.58	.59
Grco	.37	.49	.35	.50	.72	.61	.50	.52	.51
Idfc	.41	.53	.34	.57	.74	.64	.50	.52	.52
Acfc	.39	.50	.33	.56	.71	.64	.50	.51	.51
Medium text									
	ELBA							Dictation	
	1	2	3	4	5	6	7	E	D
Exact	.49	.36	.36	.56	.58	.55	.51	.58	.58
Semac	.57	.47	.43	.70	.65	.63	.59	.69	.69
Grco	.59	.36	.37	.66	.62	.61	.55	.62	.66
Idfc	.42	.33	.44	.52	.54	.49	.49	.51	.56
Acfc	.43	.29	.38	.52	.53	.48	.50	.51	.56
Easy text									
	ELBA							Dictation	
	1	2	3	4	5	6	7	E	D
Exact	.47	.47	.29	.63	.58	.55	.47	.70	.58
Semac	.55	.53	.45	.70	.69	.61	.58	.76	.67
Groo	.51	.44	.42	.65	.68	.59	.57	.74	.66
Idfc	.43	.43	.35	.57	.54	.46	.47	.64	.58
Acfc	.43	.39	.35	.58	.56	.48	.49	.69	.62

It might be supposed that grammatical scoring procedures would relate more to grammar subtests than to reading comprehension tests and the reverse for the exact word procedure. However, this did not happen. All scoring procedures correlated most highly with Grammar (test 5), on the difficult text, and either Grammar or Listening Comprehension (test 4) on the other two texts. The lowest correlations in all cases were with the segmental listening tests, but reading comprehension (7) also correlated uniformly low for all procedures. In virtually every case, the SEMAC correlated most closely with each subtest of the ELBA, regardless of text difficulty. Only rarely did the exact word procedure correlate better with the subtests than the GRCO procedure. When the dictation tests are also considered, the SEMAC again correlates higher with both tests than any other procedure. There is, however, no evidence that any one scoring procedure relates consistently more to dictation than to ELBA or one of its subtests. Moreover, as reported separately (Alderson, 1978) the results of several factor analyses did not contribute evidence to the thesis that different scoring procedures measure different language skills, since the factors that emerged were more or less equivalent across scoring procedures.

Finally the non-native speakers' performance was compared with that of the native speakers. In EFL proficiency testing, it is usually assumed that native speakers will do uniformly well on a test, whilst non-native speakers will be discriminated according to their proficiency and, in fact, t-tests on mean scores showed native speakers to be always significantly better at the close task, regardless of scoring procedure. However, the differences between means were increased by using the SEMAC procedure, as the following table shows.

TABLE 8

	Exact	Semac	Grco	Idfc	Acfc
Difficult text					
D06	4.8	8.3	6.1	3.8	4.4
D08	5.9	12.9	13.1	10.9	11.1
D10	5.3	12.3	10.5	9.9	10.1
D12	5.7	10.2	7.6	8.1	8.2
Medium text					
M06	5.4	8.2	6.5	4.4	4.7
M08	5.0	8.2	6.2	3.8	3.3
M10	6.4	10.1	7.7	5.4	5.1
M12	8.0	10.6	7.5	5.3	5.5
Easy text					
E06	4.0	3.6	2.5	1.3	1.0
E08	5.1	4.6	3.8	2.7	2.4
E10	1.8	3.2	3.0	1.9	1.9
E12	3.4	4.7	3.1	2.5	2.0

The SEMAC on both difficult and medium texts results in much greater differences than any of the other procedures. It also resulted in somewhat less overlap in performance between the two groups than the exact word procedure. In other words, if what is required of a cloze test of EFL proficiency is that it maximise the discrimination between native and non-native speakers, then the most appropriate scoring procedure is the semantically acceptable procedure.

3. Summary and discussion

To summarise the results of this study, it was found that different scoring procedures nearly always result in significantly different mean scores. Previous findings were confirmed that the exact word and the semantically acceptable procedures showed a high intercorrelation. Nevertheless, all scoring procedures showed a high level of intercorrelation, with the semantically acceptable procedure relating as much to the grammatical procedures as to the exact word procedure. The semantically acceptable procedure reduced the effect of changes in deletion frequency on mean test scores. It also gave the best item discrimination and item facility indices, and produced the highest and most consistent reliability. The semantically acceptable procedure also gave the highest and most consistent correlation coefficients with the English Language Battery, both with the total score, and with the subtests. Thus, the semantically acceptable procedure would appear to produce the best measure of EFL proficiency. There was, however, no evidence from the correlations and the factor analyses that different scoring procedures consistently measure different abilities. They seem to measure the same abilities, with greater or less efficiency. Moreover, the semantically acceptable procedure maximised the distinction between native and non-native speakers of English.

The results show that the existence of a high intercorrelation between scoring procedures is not a sufficient criterion to judge one cloze scoring procedure against another, that other evidence needs to be taken into account, and that, indeed, the other evidence may be more important and lead to different conclusions. The sum of the evidence as presented here is that the semantically acceptable scoring procedure is the most suitable for use with cloze tests. Apart from the practical consequences in the use of cloze tests, what are the implications of this finding?

It is possible to argue that the exact word procedure is the best measure of sensitivity to discourse constraints, and that therefore the semantically acceptable procedure is less sensitive to such constraints. However, no evidence was found that different procedures measure different abilities or sensitivities, and both the exact word procedure and the gramatically correct procedure related to the semantically acceptable at about the same level. We must remember, however, that we are dealing with the pseudo-random cloze procedure, in which many items deleted—function words, for instance—are usually constrained by the immediate environment, and are frequently only replaceable by the original word deleted. In a cloze test as a whole, therefore, the effect of the difference between the exact and the semantically acceptable procedures will be considerably reduced by these two features. The difference in context sensitivity, if any, will have no effect for immediately constrained items and there will in many cases only be one acceptable restoration. This suggests that cloze tests are relatively insensitive to discourse constraints, and that they are measures of somewhat lower-order skills. This suggestion is to some extent confirmed by the pattern of correlations with the English Language Battery, where the cloze related more closely to Grammar and Vocabulary than to Reading Comprehension or even the dictation. The pseudo-random cloze test appears to be a useful measure of core proficiency in English as a Foreign Language, although the variation observed in the performance of different tests means that one needs to validate each cloze test—the cloze procedure does not produce an automatically valid measure of EFL proficiency. This variation in the validity of the test can, however, be reduced by using the semantically acceptable scoring procedure.

Testing in an EFL/ESP Context

George E. Scholz and Celeste M. Scholz

Institut National d'Electricité et d'Electronique
Boumerdes, Algeria

There has been much discussion concerning the level of language proficiency at which ESP (English for Special Purposes) can most efficiently and effectively be taught to non-native speakers.

In an effort to find answers to this question, instructors at the National Institute of Electricity and Electronics in Boumerdes, Algeria, administered eight tests to 50 students who had completed a sixteen week intensive EFL program. Four of the tests were of EFL content: 1) the grammar sections of the Michigan Proficiency Exam (A and B), 2) a 100-item multiple-choice listening comprehension exam, 3) 3 short EFL cloze passages and 4) 2 EFL dictations. Four tests were of ESP content, with passages submitted by the technical faculty and produced in a testing format by the EFL staff: 1) a fifty-item multiple-choice technical grammar test, 2) 8 reading passages, 3) 3 short ESP cloze passages and 4) 2 ESP dictations. All results were correlated and factor-analyzed.

It was found that tests of EFL content correlated significantly with tests of ESP content, indicating similarity of students' scores, regardless of content. Furthermore, the factor analysis revealed all tests to be similar in that they assessed general language ability.

The following semester CELT Structure and Listening Tests were administered and the learners' final technical subject scores were collected. The scores were related to the students' previous EFL/ESP tests by regression analysis to determine the predictability of a learner's future technical performance. It was found that EFL tests predicted a student's EFL performance slightly better than the ESP tests, while the integrative tests of cloze and dictation appeared to be better indicators of a learner's ability to succeed in technical subjects.

English for Specific Purposes (ESP) enjoys a great prestige in ESL/FL. ESP rests on "a reputation for relative high success rates . . . compared with conventional teaching of English as a foreign language." (Strevens 1977c:3) In terms of establishing a coherent ESP curriculum with a learner-centered approach, impressive work has been done in identifying the learner's linguistic/communicative needs (see, for example, Jones and Roe 1975; Mackay 1978; and Munby 1977). Ewer (1975) provides an excellent discussion of the need for trained EST (English for Science and Technology) instructors.

When evaluation is taken into account, however, very little has been done in assessing an ESP program.

> It remains a major shortcoming of ESP that very little work has been done to devise fresh methods of testing, examining and assessment that match the new courses of

> training. ESP teachers combine in rejecting as unsuitable all of the many existing tests in ESL. (British Council 1976). (Strevens 1977b:129)

The paucity of adequate evaluation of both learner and curriculum presents a serious drawback to ESP. If there is no evaluation of the teaching/learning to provide feedback and subsequent adjustment of the curriculum, then the curriculum is significantly handicapped, a fact which may be a source of frustration to both the instructor and the learner.

The assumptions of the teaching/learning of ESP/EST are two: 1) The teaching/learning of EST/ESP involves both science and language. "Assumed shared knowledge, . . . presupposition. . . . affects surface syntax in EST texts so drastically that language and subject matter cannot be discussed separately when the focus is on discourse." (Selinker and Trimble 1974:83) 2) ESP/EST follows a period of intensive general English in which the learner must gain a considerable command of English. This can be attested to by a 500+ score on the TOEFL or a similar proof of proficiency of the learner before studying ESP/EST. In addition, many ESP/EST programs are designed for the post-secondary (tertiary) level of education, a time when the learners have completed many years of English.

Given these assumptions, what can an EFL/SL teacher, transformed into an ESP/EST instructor, do when his learners do not meet the proficiency criteria of assumption 2? What should the instructor do when the learners do not have a grasp of the science/technology being taught in the ESP class, let alone the language? Finally, questions within the scope of this paper are:

1) What evaluative instruments should be used?
2) What subject matter should the tests contain (general EFL or ESP)?
3) What do those instruments measure—science or language?
4) Which instrument provides a reasonable indicator of a learner's future performance in science and technology?
5) May ESP be introduced at an earlier level of a learner's English proficiency than has previously been acknowledged?

The lack of evaluative instruments in EST/ESP may be due to the controversy concerning the relationship of scientific English to general English and to the teaching of subject-matter as well as English in ESP/EST. The ESP/EST student is involved with "learning language and understanding science at one and the same time." (Boyd and Boyd 1978:25) This article reflects the effort to gain insight into the EFL/EST controversy as well as to provide satisfactory English evaluatory instruments in an ESP context. "The purpose of testing is always to render information to aid in making intelligent decisions about possible courses of action." (Carroll 1972:314) To achieve these goals, the testing in this context has been the instrument of research.

The EFL/ESP controversy listed above is further compounded when current testing research is involved. Basically, the controversy in testing research centers around the conceptions of language and language learning. The criticism directed against EFL teaching/learning is that it does not expand beyond the sentence level, i.e., it does not deal with the relationship of a particular sentence to a particular piece of discourse. This focus of instruction is the heritage of structural linguistics, based upon the premise that language learning and ability may be divided into separate (discrete) skills or components. "Discrete point analysis necessarily breaks the elements of language apart and tries to teach them separately . . . with little or no attention to the way those elements interact in a larger context of communication." (Oller 1979b) Tests based on this premise may be called discrete-point tests, which usually employ a multiple-choice format. "The most serious disadvantage of discrete-point tests in general is that they fail (in most cases) to reflect actual language usage." (Oller 1973a: 185)

The other overlapping but distinct view of language and language learning holds that "to teach a language is to teach a student to communicate in real-life situations." (Oller 1973a:185) Placing a stress on communicative rather than discrete skill competence is an accurate contemporary view of the English language teaching field. Tests of language use in meaningful contexts are "integrative" (Carroll 1972) or "pragmatic" tests (Oller 1979b).

Cloze procedure and dictation are two examples of integrative tests. As natural language is redundant, integrative/pragmatic tests exploit redundancy in a meaningful context. A cloze test reduces redundancy by a mechanical deletion of every nth word, while a dictation provides reduced redundancy via distortion. Both tests challenge the learner's internalized grammar or underlying competence of a given language. Although cloze and dictation have generally been approved as tests of reading and listening respectively, both tests have been advocated as measures of a learner's overall language proficiency (see Aitken 1977, Oller 1972a).

1. Working Hypotheses

As this study was exploratory in nature, a series of working hypotheses were formed on the assumption that those hypotheses could lead to progress in ESP evaluation.

The working hypotheses (WH) were as follows:

WH-1: There is a significant difference between various tests of English ability employing EFL and ESP content.

WH-2: Tests of English language ability employing ESP content measure a learner's science ability, not language ability.

WH-3: Tests of English language ability employing EFL content serve as better indicators of a learner's future EFL performance than tests having an ESP content.

WH-4: Tests of English language ability using ESP content serve as better indicators of a learner's future technical performance than English tests of EFL content.

WH-5: Integrative tests of English language using ESP serve as better indicators of a learner's future technical performance than discrete-point tests.

2. Setting

The tests developed for this study were administered to students at the Institut National d'Electricité et d'Electronique (INELEC) in Boumerdes, Algeria, in February, 1977. INELEC is an Algerian institution unique in its attempt to carry on an entire program in English, which is considered a foreign language, rather than French, the primary language of higher instruction. The students selected as subjects for this study were those who had completed their first semester at INELEC. The first semester consisted of a 16-week intensive EFL session, totaling 480 hours. Included in the intensive semester were courses on the English of Mathematics (80 hours) and the English of Tools (80 hours). In addition, after eight weeks the students were given an introduction to electricity (40 hours) and technical drawing (40 hours), courses which were taught by technical instructors in English.

This setting proved interesting for five reasons: 1) Almost all subjects had either Arabic or Arabic/Kabylie Berber as a first language, with French as a second language. The only exception was a student of Algerian parentage raised in France, who spoke French as a first language and Arabic as a second language. 2) All students had a relatively similar education in primary and secondary institutions. 3) All subjects had some prior English language experience, usually taught by non-native speakers with a traditional grammar orientation. Subjects could talk about grammar, but not use English for communicative purposes. Entrance tests classified students at the false beginner or low intermediate level. 4) Students entering INELEC were selected in terms of scientific and mathematical ability but not English ability. 5) None of the subjects had had previous training in electrical technology or engineering.

3. Subjects

Fifty students who had completed their first and second semester in Electrical Technology were selected for this study. Although 62 students had taken the English tests at the end of the first semester, subjects were dropped who were not passed to the second semester, who changed fields of study or did not complete the battery of tests.

4. Tests Instruments

(see Figure 1)

All the English tests used may be classified into the categories of grammar, listening and reading. With the exception of the Michigan Test of Language Proficiency (MTELP)-Grammar section (A and B-Revised) and the teacher constructed listening comprehension test, all of the tests may be found in Appendix A. The following is a description of the eight tests which were administered.

Figure 1
Tests

Skill/Focus Tested:	Content EFL	Items	Type	Content ESP/EST	Items	Type
Grammar:	Mich A+B	80	discrete-point	Tech Grammar*	50	discrete-point
Listening:	Listening	100	discrete-point			
	Dictation*	301	integrative	Dictation*	242	integrative
Reading:	Cloze*	66	integrative	Cloze*	79	integrative
				Reading Comprehension*	32	discrete-point

* These tests may be found in Appendix A.

4.1 EFL.

1) *MTELP-Grammar* (80 items) The grammar section of the MTELP, Form A (1961) and Form B-Revised (1965) were used.

2) *Listening Comprehension* (100 items) The listening comprehension was teacher constructed. Using a multiple-choice format, utterances required either an appropriate response or paraphrase. This type of listening test was similar to most standardized tests of listening comprehension.

3) *Cloze Procedure* (66 items) A set of three cloze passages with seventh word deletions were used. The passages were selected from grade school and junior high readings for native English speakers. Scoring was based on the exact and acceptable word method.

4) *Dictation* (310 items) Two dictations, taken from grade school—junior high reading, were administered. Scoring was based on the exact word method, one point for each correct word, inclusive of minor spelling errors.

4.2 ESP/EST. In developing the EST tests, the technical faculty was asked to submit two passages from their own coursework which they felt the subjects would be able to understand in terms of language and not specific scientific/technical concepts. From those passages received, items were selected to construct a discrete-point grammar test, two cloze passages and two dictations.

1) *Technical Grammar* (50 items) This test was constructed by the English teachers who compiled a list of structures taught in the intensive semester and then matching a technical sentence carrying the item. The structure to be tested was then omitted and placed in a multiple-choice format with distractors created by the English instructors. Due to a limited number of technical passages received by the English staff, a few items in the grammar test were solely the creation of the English faculty. Those items, however, contained technical classroom content.

2) *ESP/EST Reading Comprehension* (32 items) Eight short reading passages, submitted by an English of Tools instructor, were administered. Each passage

was followed by four comprehension questions in a multiple-choice format.

3) *ESP/EST Cloze Procedure* (79 items) Two passages were selected from those submitted by the technical faculty and were administered with a seventh word deletion. Another passage, submitted by an English of Tools instructor, was included. Scoring was by the exact and acceptable word method.

4) *ESP/EST Dictation* (242 items) Two passages were selected from those submitted by the technical faculty. Scoring was based on one point for each correct word, inclusive of minor spelling errors.

5. Procedure

All subjects were given various tests at different times during the examination week of the first semester. The logistics of providing one test at a time to all subjects proved impossible. All tests and answers were collected at the end of each test period. All subjects had practice in cloze procedure and dictation prior to the examination.

At the end of the subjects' second semester, final scores for the subjects' technical courses were collected. The technical courses were Mathematics, Technical Drawing and D-C Circuitry. The subjects were also given the Comprehensive English Language Test (CELT): Listening (Form L-A, 1970) and Structure.

6. Statistical Procedure

All results of the first semester were correlated and factor analyzed, using Pearson product-moment correlation (r) and principal component analysis. The use of r and factor analysis permitted testing of WH-1 and WH-2.

A correlation indicates the associative relationship (if any) between two variables. The correlation squared (r^2) indicates the amount of variance shared by two variables. "Factor analysis is one of the statistical techniques for examining . . . patterns of correlation." (Oller and Hinofotis 1976:1) Highly correlated variables form (or load on) a factor. It may be hypothesized that the variables on a factor share a common or underlying source. What is of importance is the amount of loading of each variable on a factor, which indicates the factor's importance to a variable. Using a principal components solution, if a general or common factor, G, indicating similarity between variables, is to be accepted, the product between two variable loadings (predicted r) should equal the actual correlation between them. The remaining (residual) variance between the two variables should be near or at zero.

To test WH-3, WH-4 and WH-5 the scores of the eight tests administered at the end of the first semester were correlated with the technical and English test scores of the second semester. With each of the scores of the second semester serving as dependent variables and each of the English tests of the first semester serving as independent variables, a simple linear regression was calculated to determine which one of the eight English tests would be the best predictor for each of the dependent variables.

Given the correlation between two variables, simple regression analysis provides the best prediction possible. (Kerlinger 1973:604) From a one-way analysis of variance, the F-ratio may be found indicating the statistical significance of the regression, of predicting Y (the dependent variable) from X (the independent variable).

7. Results and Discussion

The mean, standard deviation, standard error of measurement and reliability estimates of all tests may be found in Table 1. The correlation matrix for

TABLE 1

I. Semester I

Test	Items	Mean	Standard Deviation	Standard Error of Measurement	Reliability (r_{11})*
Mich Grammar	80	41.84	7.37	4.42	.64
Listening	100	64.88	9.25	4.71	.74
Cloze EFL	66	39.04	8.88	4.87	.81
Dictation EFL	301	130.22	31.56	8.41	.93
Total EFL	547	272.40	51.37	11.50	.95
Tech Grammar	50	29.38	6.71	3.38	.75
Reading Comp EST	32	17.96	4.83	2.72	.68
Cloze EST	79	48.70	10.12	4.20	.83
Dictation EST	242	136.48	27.30	7.53	.92
Total EST	403	232.18	42.14	9.71	.95
II. Semester II					
English					
CELT Structure	75	41.84	7.37	4.23	.67
CELT Listening	50	28.44	5.36	3.47	.58
CELT Total	125	63.48	11.02	5.51	.75
Technical					
Math	100	53.98	18.08	4.78	.93
DC Circuits	100	58.74	9.57	4.88	.74
Technical Drawing	100	86.18	16.61	2.88	.97
Total Technology	300	198.96	37.54	7.51	.96

* Kuder-Richardson reliability approximation taken from *Introduction to Research in Education*, D. Ary, L. Jacobs, A. Razavieh. New York: Holt, Rhinehart & Winston, 1972, p. 207-208.

TABLE 2a

Correlation Matrix of Four EFL and Four ESP/EST tests;
Observed r above diagonal and Common Variance (r^2) below diagonal.

	1	2	3	4	5	6	7	8	9	10
1 Mich Grammar		.62	.60	.62	.79	.79	.53	.63	.62	.73
2 Listening	.38		.53	.44	.64	.57	.63	.51	.64	.70
3 Cloze EFL	.36	.28		.43 (.002)	.57	.66	.50	.69	.51	.65
4 Dictation EFL	.38	.19	.18		.91	.57	.30 (.033)	.68	.77	.77
5 Total EFL	.62	.41	.32	.83		.76	.50	.78	.81	.88
6 Tech Grammar	.62	.32	.44	.32	.58		.58	.72	.61	.78
7 Reading Comp EST	.28	.40	.25	.09	.25	.34		.47	.43 (.002)	.60
8 Cloze EST	.40	.26	.48	.46	.61	.52	.22		.66	.83
9 Dictation EST	.38	.41	.26	.59	.66	.37	.18	.44		.95
10 Total EST	.53	.49	.42	.59	.77	.61	.36	.69	.90	

p is less than .001 unless indicated.

the EFL/ESP tests may be found in Table 2a. Table 2b contains the correlation matrix of technical scores. Table 3a contains the factor analysis and calculated loading of the EFL/ESP tests. Table 3b contains a comparison of predicted correlations from the factor loading with the actual correlations. Table 3c contains the remaining variance not accounted for by the factor loadings.

The correlations found in Table 2a reveal that there is a good deal of shared variance between the Total EFL and Total ESP tests ($r = .88$, $r^2 = .77$). This high correlation indicates that there is little significance attributed to content with regard to a subject's score. It appears that EFL tests of one skill generally correlate more highly with ESP tests of the same skill than other tests

TABLE 2b

Correlation Matrix of Technical Scores
Observed r above diagonal and Common
Variance (r^2) below diagonal. ($N = 50$)

	1	2	3	4
1 Math		.61	.60	.90
2 DC Circuits	.37		.45	.75
3 Technical Draw	.36	.20		.85
4 Total Tech	.81	.56	.72	

p is less than .001.

TABLE 3a

I. Principal Factor Solution (with iterations) for the four EFL tests and four EST tests ($N = 50$ Algerian subjects).

Tests	Loadings on G Factor*	h^2
Michigan Grammar	.83	.69
Listening	.73	.53
Cloze EFL	.73	.53
Dictation EFL	.72	.52
Technical Grammar	.86	.74
Reading Comprehension EST	.63	.40
Cloze EST	.82	.67
Dictation EST	.80	.64
	Eigen value	4.72

* Accounts for 100% of the total variance in the factor matrix.

II. Number of Factors in Varimax Rotation less than two, so Rotation Bypassed.

TABLE 3b

Correlation Matrix (above the diagonal) and Predicted Correlations Derived from Respective Products of Loadings on G (below diagonal).

	1	2	3	4	5	6	7	8
1 Mich Gram		.79	.53	.62	.60	.63	.62	.62
2 Tech Grammar	.71		.58	.57	.66	.72	.61	.57
3 Reading Comp EST	.52	.54		.63	.50	.47	.43	.30
4 Listening	.61	.63	.46		.53	.51	.64	.44
5 Cloze EFL	.61	.63	.46	.53		.69	.51	.43
6 Cloze EST	.68	.71	.52	.60	.60		.66	.68
7 Dictation EST	.66	.69	.50	.58	.58	.66		.77
8 Dictation EFL	.60	.62	.45	.53	.53	.59	.58	

TABLE 3c

Residual Matrix with G Loading Partialed Out (mean of absolute values = .0625): Observed r minus Product of Loading on G.

	1	2	3	4	5	6	7	8
1 Mich Gram		.08	.01	.01	−.01	−.05	−.04	.02
2 Tech Grammar			.04	−.06	.03	.01	−.08	−.05
3 Reading Comp EST				.17	.04	−.05	−.07	−.15
4 Listening					.00	−.09	.06	−.09
5 Cloze EFL						.09	−.07	−.10
6 Cloze EST							.00	.09
7 Dictation EST								.19
8 Dictation EFL								

of different skills or content. The Michigan Grammar test correlates most closely with the Technical Grammar test ($r = .79$). Listening Comprehension correlates highly with Dictation EST ($r = .64$). Cloze EFL correlates most highly with Cloze EST ($r = .69$). The test with the poorest correlations appears to be the Reading Comprehension-EST test, which also has a low reliability.

The factor analysis reveals a single unitary factor accounting for 100% of the variance in the total matrix. All the variables may then be hypothesized as tests of general language ability. All variables load highly on the G factor with the exception of the Reading Comprehension test ($h^2 = .40$). The residual variance (Table 3c) indicates that only a small amount of variance remains unaccounted for by the G factor.

The factor analysis does not reveal different skill areas with unique variance. If this were the case, various factors would be produced, corresponding to a skill. In this study, the tests of grammar, listening and reading would produce three different factors. Furthermore, if there were also a difference between language content and skill, there would be six factors: one for EFL Grammar, one for Tech Grammar, and so on. If the EST tests measured science and not language ability, then there would be two factors: one for science, containing the ESP tests and one for language, containing the EFL tests. The data, however, do not support any of these conditions. Rather, they appear to support the *unitary competence hypothesis* (Oller and Hinofotis 1976; Scholz, Hendricks, et al. 1979) that "The components of language competence, whatever they may be, may function more or less similarly in any language-based task." (Oller and Hinofotis 1976:2)

Table 4a contains the correlations and F-ratios of the regression analysis of the English CELT tests on the eight EFL/ESP tests. Table 4b contains the correlations and F-ratios of the regression analysis of the technical scores on the English tests.

All regressions of the English tests were significant at the .01 level with the exception of the CELT Structure Test on Cloze-EST ($p < .05$). Tests of grammar correlated most highly with the CELT Structure test, (Table 4a-1) although Michigan Grammar correlated slightly higher than the Tech Grammar. Tests of listening correlated most highly with the CELT Listening test, (Table

TABLE 4a

Linear Regression Analysis of English Tests (N=50)

4a-1 Dependent Variable: CELT Structure

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Mich Grammar	.69	.48	1096	1326	39.67*
2 Tech Grammar	.69	.47	1014	1306	37.27*
3 Listening Comp	.52	.27	504	1818	13.31*
4 Dictation EST	.48	.23	639	1880	16.31*
5 Cloze EFL	.46	.21	510	1556	15.73*
6 Dictation EFL	.45	.20	640	1948	15.77*
7 Cloze EST	.41	.16	185	1374	6.46**
8 Reading Comp EST	.34	.11	965	2216	20.90*

4a-2 Dependent Variable: CELT Listening

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F*
1 Listening Comp	.68	.47	583	784	35.69
2 Dictation EST	.63	.40	456	858	25.21
3 Dictation EFL	.62	.39	556	860	31.03
4 Cloze EST	.52	.28	651	1056	29.59
5 Mich Grammar	.50	.25	262	1040	12.09
6 Tech Grammar	.48	.23	375	1100	16.36
7 Reading Comp EST	.43	.18	401	1146	16.80
8 Cloze EFL	.40	.16	342	1186	13.84

* p is less than .01

** p is less than .05

4a-3 Dependent Variable: CELT Total

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F*
1 Mich Grammar	.69	.47	3878	3152	59.06
2 Tech Grammar	.67	.45	2158	3438	30.13
3 Listening Comp	.66	.44	2829	3392	40.03
4 Dictation EST	.62	.38	2246	3700	29.14
5 Dictation EFL	.59	.35	2467	3770	31.41
6 Cloze EST	.52	.27	1441	4372	15.82
7 Cloze EFL	.49	.24	1940	4616	20.17
8 Reading Comp EST	.42	.18	952	5204	8.78

* p is less than .01

TABLE 4b

Linear Regression Analysis of Technical Course Grades (N=50)

4b-1 Dependent Variable: Math

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	.23	.05	823	14,764	2.68
2 Reading Comp EST	.09	—+	−406	15,926	——
3 Cloze EFL	.07	—+	194	15,994	——
4 Dictation EFL	.07	—+	−250	15,878	——
5 Dictation EST	.07	—+	180	15,338	——
6 Mich Grammar	.06	—+	69	15,824	——
7 Tech Grammar	.05	—+	−57	16,096	——
8 Listening Comp	.01	—+	108	16,010	——

+ less than .01

4a-2) although the EFL Listening Comprehension correlated slightly higher than Dictation-EST. In regard to the Total CELT scores, (Table 4a-3) both discrete-point and integrative tests, regardless of EFL or ESP content, appeared to serve as indicators of a subject's future English ability

With respect to the regression analysis of the technical scores on the eight English tests (Table 4b), the picture is not so clear. The Cloze-EST test correlates the highest of all the English tests with the Technical scores, with the exception of D-C Circuitry. However, the regression analysis was not significant at the .05 level. Cloze and dictation, integrative tests, served as slightly better indicators than the discrete-point tests of grammar.

One problem with the regression analysis may be that the technical scores were not accurate representations of the subjects' technical performance. Other factors besides technical ability may have been taken into account in the final technical score calculations. The assumption that knowledge of English would have no bearing at all in an entirely English academic program seems tenuous.

To explore further the relationship of EFL/ESP to future technical performance the subjects with the highest scores in technology were examined. The criterion was that the subject had to have scored one standard deviation or better on three or more of the technical scores.[1] Nine subjects met the criterion. The results may be found in Table 5. Although very few of the regressions are statistically significant at the .05 level, the correlations are much larger. Cloze-EST still maintains the highest correlation with the technical scores. Both grammar tests and the Listening Comprehension test do not appear to predict as well as some of the integrative tests of cloze and dictation. In terms of technical performance, it appears that in the case of the regression of all subjects (Table 4b-4) and of the top subjects (Table 5d) Cloze-EST and Dictation-EFL are the best predictors of the eight English tests.

8. Conclusions and Recommendations

Tests containing either EFL or ESP measure general language ability. The ESP/EST tests used in this study assess language and not science. EFL tests are slightly better indicators of a learner's future EFL performance than ESP tests.

In terms of predicting technical scores, more research is needed to determine EFL/ESP tests that are able to indicate future technical performance at a statistically significant level. This study speculates that integrative tests may serve as better indicators of technical performance than discrete-point tests. An integrative/pragmatic test in an EST context is probably more valid than a discrete-point test, as "the validity of the test can be established not solely on the basis of whether it appears to involve a good sample of the English language but *more on the basis of whether it predicts success in the learning tasks and social situations to which the examinees will be exposed.* (emphasis mine)" (Carroll 1972:319)

---

[1] One exception was Technical Drawing, where the criterion was a score of 95 or better.

4b-2 Dependent Variable: D-C Circuitry

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Listening Comp	−.25	.06	284	4214	3.23
2 Cloze EST	.07	−+	256	4438	2.77
3 Cloze EFL	.06	−+	143	4466	1.54
4 Reading Comp EST	−.03	−+	−107	4496	——
5 Mich Grammar	−.03	−+	361	4446	3.90
6 Tech Grammar	−.03	−+	361	4454	3.89
7 Dictation EST	−.03	−+	478	4418	5.19**
8 Dictation EFL	−.02	−+	1180	4438	12.76*

4b-3 Dependent Variable: Technical Drawing

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	.31	.09	366	12,284	1.43
2 Dictation EFL	.27	.07	1526	12,336	5.94**
3 Cloze EFL	.20	.04	1400	13,052	5.15**
4 Mich Grammar	.18	.03	115	12,990	——
5 Dictation EST	.11	.01	10	20,558	——
6 Tech Grammar	.07	−+	69	13,526	——
7 Reading Comp EST	−.08	−+	103	13,446	——
8 Listening Comp	−.04	−+	27	13,360	——

* p is less than .01
** p is less than .05
+ less than .01

4b-4 Dependent Variable: Total Technology

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	.27	.07	1259	62,814	.96
2 Dictation EFL	.15	.02	1195	66,976	.86
3 Cloze EFL	.14	.02	1436	67,198	1.03
4 Mich Grammar	.11	.01	1957	68,446	1.37
5 Dictation EST	.08	−+	−1216	68,232	——
6 Listening Comp	−.07	−+	392	68,210	——
7 Tech Grammar	.05	−+	605	68,895	——
8 Reading Comp EST	−.0001	−+	796	69,068	——

+ less than .01

TABLE 5

Linear Regression Analysis of Technical Scores of Top Subjects in Technical Courses (N=9)

5a Dependent Variable: Math

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	.61	.37	240	428	3.93
2 Dictation EFL	.55	.31	202	452	3.13
3 Reading Comp EST	.37	.14	249	552	3.16
4 Dictation EST	.31	.10	368	574	4.49
5 Tech Grammar	.30	.09	52	606	——
6 Listening Comp	.24	.06	34	578	——
7 Mich Grammar	−.23	.05	−123	600	——
8 Cloze EFL	−.006	0	312	644	3.39

The integrative tests of EST-Cloze and Dictation-EFL may indicate that, in a technical context, listening and reading are significant language skills. It

5b Dependent Variable: D-C Circuitry

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	.63	.39	50	84	4.17
2 Cloze EFL	.58	.34	185	102	12.70*
3 Dictation EST	.56	.31	−99	96	——
4 Reading Comp EST	.56	.31	−228	98	——
5 Tech Grammar	.36	.13	151	126	8.36**
6 Mich Grammar	.31	.10	12	128	——
7 Listening Comp	.30	.09	−101	114	——
8 Dictation EFL	.06	.03	276	150	12.88*

* p is less than .01
** p is less than .05

5c Dependent Variable: Technical Drawing

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	−.70	.49	223	34	41.91*
2 Tech Grammar	−.63	.39	−161	42	——
3 Dictation EST	−.54	.29	−211	52	28.40*
4 Listening Comp	−.50	.25	205	60	23.92*
5 Mich Grammar	−.50	.25	−171	52	——
6 Dictation EFL	−.22	.05	195	70	19.50*
7 Cloze EFL	−.21	.04	4	68	——
8 Reading Comp EST	−.05	−+	−380	66	——

5d Dependent Variable: Total Technology

Source (in order of correlation)	r	r^2	Sum Squares-Regression (df=1)	Sum Squares-Residual (df=48)	F
1 Cloze EST	.51	.26	159	384	2.90
2 Reading Comp EST	.42	.18	109	512	1.49
3 Dictation EFL	.40	.16	89	506	1.25
4 Dictation EST	.29	.09	47	564	——
5 Listening	.28	.08	−430	550	——
6 Mich Grammar	−.21	.05	−444	582	——
7 Tech Grammar	.20	.04	−456	602	——
8 Cloze EFL	.07	−+	5	616	——

* p is less than .01
+ less than. 01

may be that while the learner may have to read his technical books, the technical instructor simplifies his scientific information into everyday English. As Michael Collins of the EFL staff of the University of Petroleum and Minerals asserts after observing the language of science lectures at his university,

> The need for communication forces the science teacher to explain difficult or unfamiliar terms and concepts by reference to everyday examples in everyday language and this is the kind of language he uses most of the time. (Boyd and Boyd 1978:25).

Needless to say, the development of reliable and valid tests for learners in an ESP context remains necessary. It is hoped that the statistical analysis and ideas in this paper may help this development.

Appendix A

TECHNICAL GRAMMAR

1. "You understand electronic theory very well."
"I should. I ____________ it ever since I began INELEC."

a) haven't studied
b) had studied
c) have been studying
d) had been studying

2. The drawing must convey the information ____________ or the part will not be made correctly.

a) accurate and complete
b) accuracy
c) accurately and completely
d) as accurate and complete

3. You ____________ what happens when safety precautions aren't followed.

a) have ever seen
b) have never seen
c) are usually seen
d) can be seen

4. The charge which might exist could be ____________ positive or negative depending on which material gives up electrons more easily.

a) more
b) either
c) neither
d) such

5. "Carefully inspect your circuit, checking the polarity of the leads." He said ____________.

a) to inspect your circuit carefully
b) inspect your circuit carefully
c) you inspect your circuit carefully
d) not inspect your circuit carefully

6. "Don't touch that wire!" The wire is ____________ dangerous to touch.

a) much
b) very
c) such
d) too

7. The word electronics derives from the electron. Electronics can ____________ defined to include all applications of electricity flowing in a vacuum.

a) for
b) be
c) have
d) been

TECHNICAL GRAMMAR

8. Ahmed ____________ seen a voltmeter before he came to INELEC.

a) had not
b) has never
c) hadn't been
d) hasn't

9. When working on printed circuits, too much heat ____________ soften the plastic form and cause damage.

a) has
b) can
c) must
d) has to

10. Unless the student has a clear mental picture of a letter, he ____________ make the letter correctly.

a) should
b) shouldn't
c) can't
d) can

11. Lithium doesn't have six electrons in its second shell, and ____________ does beryllium.

a) either
b) also
c) neither
d) so

12. Inside, a glass rod supports a frame of wires ____________ holds up a very thin, coiled wire, called a filament.

a) who
b) whose
c) that
d) these

13. Rosin-core flux is the solder most often used. Scientists, ____________ though that flux is not a substitute for cleaning the metals to be soldered.

a) have noted
b) have been noted
c) has noted
d) have being noted

14. The secret in soldering is to heat the joint, not the solder. When the joint is ____________ to melt the solder, it forms a cover without any air spaces.

a) too cold
b) enough hot

Appendix A (Continued)

TECHNICAL GRAMMAR (continued)

c) hot enough
d) very hot

15. In today's mass production techniques, thousands of identical parts ________ made from one drawing.

a) should be
b) can't be
c) must be
d) may be

16. If the wire ________ broken at any point, electrons would build up at the end of the wire that is connected to the negative side of the battery.

a) were
b) is
c) has
d) had

17. "How can the current be measured?" The instructor asked ________.

a) how the current could be measured
b) if the current could be measured
c) how could the current be measured
d) if the current can be measured

18. If two or more views are required, then the question of logical arrangement of the views arises. ________ drawings have three views of the object.

a) most
b) least
c) few of
d) little

19. Good drafting letters are well-proportioned. For example, the cross bar of the letter A ________ at one-third of the height of the letter.

a) should be
b) can't have
c) won't be
d) might be

20. The legibility of the drawing often ________ how well the letters are made.

a) decides
b) is like
c) depends on
d) applies to

21. The polarities on the meter terminal must be observed to obtain a positive meter reading. If the connections were ________, the pointer would deflect in the reverse direction.

a) reverse
b) reversing
c) reverses
d) reversed

22. The student ________ put down the compass when the technical drawing teacher told him to stop.

a) has already
b) has just
c) had just
d) has never

23. "What kind of meter are you going to use?"
"I'd like to use something ________ the one you are using."

a) that
b) like
c) such
d) as

24. The mathematical proof of the theorems of geometry are not ________ important to the draftsman ________ the application of these theorems to practical problems.

a) as ________ as
b) more ________ of
c) most ________ of
d) so ________ that

25. You ________ learned electronic theory. You are now ready to put it into practice.

a) have not
b) have yet
c) hadn't
d) have already

26. "Don't turn on the power before you close the switch." The directions said ________.

a) don't turn on the power
b) turn on the power
c) not to turn on the power
d) to not turn on the power

27. If you were to bring two magnets together with the north poles facing each other, you ________ feel a force of repulsion between the poles.

a) will

TECHNICAL GRAMMAR (continued)

b) would
c) won't
d) had

28. One student works hard in his lab class. The other student ________ works.

a) hard
b) hardly
c) completely
d) doesn't

29. Before the electric light bulb and its contents are sealed up tight, all the air ________ out and a special gas put in.

a) had been taken
b) are taken
c) is taken
d) will have been taken

30. The two types are acid flux and rosin flux. Acid flux is ________ active in cleaning metals than the other.

a) least
b) as
c) more
d) that

31. "Are you going to put on your safety glasses?"
He asked ________.

a) if were you going to put on your glasses
b) where you are going to put on your glasses
c) are you going to put on your glasses
d) if you were going to put on your glasses

32. The atoms of conductors have only 1 or 2 valence electrons. If you ________ at the atomic table, you can identify the good conductors.

a) looked
b) had looked
c) look
d) have been looking

33. Neon is a naturally stable element, and ________ is helium.

a) so
b) neither
c) either
d) such

34. A certain type of ohmmeter must be zeroed each time you change the range.

TECHNICAL GRAMMAR (continued)

When the adjustment ________ deflect the pointer all the way to zero, it usually means the dry cells must be replaced.

a) should not
b) might
c) cannot
d) had better

35. Recently man ________ a new form of power, atomic power.

a) are being discovered
b) had discovered
c) has discovered
d) has been discovered

36. "Did you measure the resistance?"
"No. I ________ to do it during the lab hour, but I didn't have time."

a) hadn't planning
b) had been planning
c) have planned
d) had been planned.

37. Voltage ________ by using a voltmeter.

a) measures
b) is measuring
c) had measured
d) is measured

38. A student would have difficulty making a letter correctly ________ a clear mental picture of the letter.

a) if he has
b) if he had
c) unless he had
d) if he doesn' have

39. Atoms with only one valence electron are the best electrical conductors, ________ copper is a good conductor.

a) because
b) so
c) as
d) but

40. The draftsman is ________ important person that his mistakes cause his company to loose money.

a) an
b) an such
c) such an
d) so

41. Electrons in the outer orbits of an atom are "free" electrons and may be

TECHNICAL GRAMMAR (continued)

easily forced from their orbits. They are attracted to the nucleus by ________ force than electrons whose orbits are near the nucleus.

a) less
b) lesser
c) greater
d) more

42. When the big block of ice that ________ Europe in the ice age began to melt, forests began to appear.

a) have covered
b) had been covered
c) had being covered
d) had covered

43. The VOM is used to measure voltage, resistance and current. It is called a Multimeter ________ it has multiple uses.

a) while
b) when
c) so
d) because

44. A little knowledge can be a dangerous thing and if something can go wrong, it ________

a) will
b) should
c) had
d) would

45. Working with electricity is ________ dangerous that the safety rules should be followed at all times.

a) so
b) very
c) too
d) such

TECHNICAL GRAMMAR (continued)

46. Technical drawings that are not clear and accurate are difficult to read; ________, a good draftsman draws cleanly and accurately.

a) when
b) while
c) because
d) therefore

47. Thomas Edison invented the light bulb in 1879. Since then, men ________ many new electrical improvements.

a) made
b) have made
c) is made
d) have been made

48. The American system place the top view above the front view. The European system, does the opposite. It places the top view ________ the front view.

a) on
b) below
c) over
d) in

49. "Why can't I use this meter?"
"Because it's ________ that doesn't work now."

a) the only one
b) only one
c) only of one
d) only the one

50. The directions say to draw an arc on either side of the triangle. The tirangle is ________ the arcs.

a) around
b) between
c) right of
d) left of

DICTATION

EST 1. A volt is the unit of electrical pressure or potential./ Voltage is measured by using a voltmeter./ Voltmeters have a high internal resistance/ and are always conected in parallel/ with a circuit or component such as a resistor./ The polarities marked on the meter terminals/ must be observed to obtain a positive meter reading./ If the connections are reversed/ the pointer will deflect in the reverse direction./ The ampere is the unit of electric current./ Current is measured by using an ammeter./ Ammeters have low internal resistance/ and are always connected in series/ with a circuit or component such as a resistor./

EST 2. Potential difference is necessary/ to produce electrical current./ The number of free electrons that can be forced/ to drift through the wire to produce the moving charge/ depends upon the amount of potential difference across the wire./ With more applied voltage the forces of attraction/ and repulsion can make more free electrons drift/ producing more

DICTATION (continued)

charge in motion./ A larger amount of charge moving with the same speed/ means a higher value of current./ Less applied voltage across the same wire/ results in a smaller amount of charge in motion/ which is a smaller value of current./ With zero potential difference across the wire/ there is no current./

EFL 1. Every school child today knows how to tell the time./ It seems hard to believe that early man/ had no way to tell time./ In order to tell when things had happened/ men had to learn to measure the passing of time./ Men learned to measure a year by the time/ that it took for the earth to go through the seasons./ They learned to measure a day/ by the time it took the sun to travel across the sky./ Long after men could tell one day from another/ they could still not measure the time within a day./ Now we can measure our days in seconds, minutes and hours./ We have clocks to tell us the time of day./ We can measure our years in days, weeks and months./ We have calendars to tell us the time of year./

EFL 2. In cities in the United States/ there are clocks in most stores, factories and other buildings./ Radio announcers give the correct time during the day./ People here think that it is important to know the time./ Most Americans have watches./ They want to do certain things at certain times./ They don't want to be late./ Time is not so important to people everywhere./ Suppose you visit a country in South America./ You would find that people living there do not like to *rush*./ If you had an appointment with someone/ the other person would probably be late./ He would not want to arrive on time./ In South America/ even the radio programs may not begin right on time./ Nor do the men on the radio think it important to announce the exact time./ In South America many people think of a clock as a machine./ They feel that a person who does everything on time/ lets a clock run his life./ They don't want a clock or any machine/ to have that much power over their lives./

CLOZE

EST 1. The VOM (Volt-Ohm-Millimeter) is the instrument used most often by the technician. It is used to measure voltage, resistance, and current; and because of this multi-usage is also called a Multimeter. In electronic shops and ________ industry, you will hear it called ________ VOM or Multimeter.

It is important ________ the student of electronics learn this ________ and how to use it correctly ________ safely. What bread and water are ________ the human, the VOM is to ________ technician. You must learn to read ________ scales, where to connect the meter ________, how to position the range and ________ switches, and how to measure voltage, ________ and current.

Voltage is an electromotive ________. It is the force from a ________ or generator that drives our radios ________ electrical appliance. Voltage is the force ________ causes current to flow. Voltage is ________ in volts.

Current is the movement ________ electrons through the wire. This movement ________ forced by the voltage. Current is ________ in amperes or milli-amperes.

Resistance ________ the opposition to current flow. It ________ measured in ohms.

EST 2. Two cases of zero potential difference and no current can be considered in order to emphasize the important fact that potential difference is needed to produce current. Assume a copper wire to be ________ itself, not connected to any voltage ________, so that there is no potential ________ across the wire. The free electrons ________ the wire can move from atom ________ atom, but this motion is random, ________ any organized drift through the wire. ________ the wire is considered as a ________, from one end to the other, ________ current is zero. As another example, ________ that the two ends of the ________ have the same potential. Then free ________ cannot move to either end, because ________ ends have the same force, and ________ is no current through the wire. ________ practical example of this case of ________ potential difference would be connecting both ________ of the wire to just one ________ of a battery. Each end of ________ wire would have the same potential ________ there would be no current. The ________, therefore, is that two connections are ________ to two points at different potentials ________ order to produce current.

EST 3. There are many kinds of tools. ________ useful kind is the screwdriver. There ________ two types of screwdriver but both ________ have the same purpose: to tighten ________ loosen screws. The type used for ________ or loosening Phillips-head screws is ________ the

CLOZE (continued)

Phillips screwdriver. The one which ______ and loosens standard screws is the ______ screwdriver. No matter which type is ______, the process is the same. It ______ very important to select the correct ______ size for the particular job you ______ are going to do. The screwdriver must ______ the screwhead exactly. If the screwdriver ______ is too small or too large ______ the head of the screw it ______ probably damage it. Once the correct ______ size has been selected, the screw ______ be tightened by turning it clockwise. ______ loosen the screw, on the other ______, turn it in a counterclockwise ______.

Wrenches, another type of tool, come ______ a variety of shapes and sizes. ______ screwdrivers, wrenches are used to make ______ tighter or looser. They are used ______ tightening nuts and bolts. One of ______ most important types of wrench is ______ at both ends and is called ______ open-end wrench for that very reason. ______ either end of this kind of ______ may be put to use, it ______ said to be reversible. The box-end ______ almost always has two sides too. ______ the open-end type, it is reversible—______ is a useful end on either ______ of the handle.

EFL 1. From here to there

Long, long ago, there was no way to tell time. There were no clocks. There were ______ calendars. There were no rulers, either. ______ did not think about how long ______ how wide anything was until they ______ to own land. Then, many kinds ______ measurements were needed.

To own land, ______ man had to know where his ______ ended and someone else's began. To ______ a house, he had to know ______ big to make it. To trade ______ goods for things he wanted, he ______ to know how much his things ______ worth. To know when to plant ______ crops, he had to have some ______ of telling time.

Little by little, ______ learned to measure to find out ______ things he needed to know.

EFL 2. Half Go Hungry

There are over two billion people in the world. Each day, around 324,000 babies are ______, and the world's population grows larger. ______ the present time, there is not ______ food for everyone. More than half ______ world's population does not get enough ______ the right kinds of food to ______ healthy. About 10,000 people in the ______ die of hunger every day.

Trying ______ feed the world's billions is a ______ task. It is a job too ______ for some nations working alone. It ______ a job for many nations of ______ world working together. With the help ______ modern science, tools, and machinery, many ______ can be helped to grow more ______.

Over sixty countries have joined together ______ form the Food and Agriculture Organization ______ the United Nations. These nations work ______ to fight hunger.

The Food and ______ Organization searches for new and better ______ to grow food. It helps fight ______ and plant diseases. It sells seed ______ farmers at low cost. It works ______ turn deserts and waste lands into ______ lands. It helps to improve the ______ of the world's peoples. It sends ______ of agriculture into those countries that ______ help.

EFL 3. Science and Technology

There is little doubt that science and technology are closely related. In fact, technology is based on ______ knowledge. Scientists are constantly discovering new ______ and gathering additional information about matter ______ energy. Putting these principles and information ______ use is technology.

Although science and ______ are usually closely related, improvements in ______ do not necessarily require a great ______ of scientific knowledge. In ancient times, ______ example, the wheel was used. The ______ of the wheel was a great ______ advance. Pulleys and windmills were invented ______ used long before the term scientist ______ into use.

Modern technology seems to ______ getting more and more dependent on ______ scientists. Scientists have been the key ______ in several examples of modern advances ______

CLOZE (continued)

technology. For example, computers, atomic reactors, ________ lasers were developed by teams of ________.

Reading Comprehension EST

The human ear can hear a wide range of sounds. This range is measured in units called decibels. The sound of a whisper has a decibel level of 20, and the sound of conversation has a decibel level of 60, and the sound of an airplane is about 140 decibels. Sound becomes noise (and can begin to damage your hearing) when it goes above 86 decibels. Of course sounds below 86 can be called noise when they interrupt other sounds and are unwanted. Doctors have found that electric guitars and other amplified instruments may produce decibel levels from 90 to 105. These levels, according to some doctors, can begin to cause deafness in the musicians and their listeners.

1. Which of these does not define noise?
    1. a sound above 86 decibels
    2. sound that is unwanted
    3. a sound that interrupts other sounds
    4. a sound with a wide range
2. What does "unit" mean in this passage?
    1. an amplified guitar
    2. a fixed amount
    3. a noise
    4. an inch
3. If something had a decibel level of 60 it would *not* be:
    1. a sound
    2. a noise
    3. a cause of deafness
    4. unwanted
4. The decibel level at which a sound can be called a noise is:
    1. not exact and absolute
    2. determined by doctors
    3. 105 decibels
    4. the decibel level of airplanes

A phonograph is a thin, flat plastic disk in the shape of a circle. Sound is recorded on a phonograph record in one long spiral groove. Sound results from the movement of waves in air. These movements, or vibrations, direct a sharp knife-like cutter that scratches a wavy groove into the plastic disk. The cutter moves in a continuous path around the disk from the edge to the center. The patterns of the waves in the groove change as the pitch and loudness of the sound change. When the disk is played on a phonograph at the same speed at which the sound was recorded, the waves in the groove make the phonograph needle and arm vibrate. An amplifier is used to make these vibrations strong enough to hear as sound.

5. A phonograph needs all of the following to make sound *except*:
    1. The record moving at the same speed that the cutter moved
    2. an amplifier
    3. a knife
    4. air
6. If the patterns of the waves in the groove do not change, the sound on the record will:
    1. be louder and higher than normal
    2. not change
    3. not be heard without an amplifier
    4. need a sharper phonograph needle to make the sound change

Reading Comprehension EST (continued)

7. The word "spiral" in this passage means:
   1. circles inside each other
   2. an unending line
   3. a cut
   4. a path that circles around in smaller and smaller circles like a coil
8. The function of the amplifier is to:
   1. create vibrations
   2. increase the strength of the vibrations
   3. change the groove pattern
   4. make the disk stronger

Matches are called safety matches when they can be lit only by rubbing them against the striking surface on the outside of the box in which they are packaged. The match "lights" because there is a reaction between the chemicals in the head of the match and the chemicals on the striking surface of the box.

The composition of the match head includes a chemical that carries oxygen, such as lead oxide; a chemical that is inflammable, such as sulfur; and substances that cause friction, such as powdered glass. Glue-like elements bind everything together.

When the head of the match is rubbed against the striking surface, friction causes heat on a small area of the head. The heat frees the oxygen from the chemical carrying it, and the oxygen joins with the sulfur to form sulfur dioxide. This causes more heat, which in turn causes more oxygen to be freed and to be combined with sulfur to make more sulfur dioxide. The chemical process is so fast that the match appears to catch fire all at once.

9. A safety match needs all of the following to light *except*:
   1. a chemical that carries oxygen
   2. coloring matter
   3. the striking surface of the box
   4. friction
10. What does "composition" mean in this passage?
    1. a written page
    2. a combination of things
    3. the shape of the match head
    4. powdered glass
11. What does "striking surface" mean in this passage?
    1. the box
    2. the outside part of the box
    3. cardboard
    4. the label
12. The purpose of this passage is to:
    1. describe how a safety match works
    2. warn against unsafe matches
    3. list the chemicals in safety matches
    4. describe how oxygen combines with sulfur

During a radio broadcast or tape playback, the level meter indicates the condition of the batteries. If the pointer is in the GOOD range, the voltage is sufficient. If it is in any other position, all the batteries should be replaced with new ones. If the batteries are weak, slow tape speed, sound distortion or low volume will occur.

13. The level meter operates only:
    1. if the batteries are weak
    2. when the set is on

Reading Comprehension EST (continued)

3. when the pointer indicates GOOD
4. when there is sound distortion

14. Slow tape speed and sound distortion are the result of:

    1. low volume
    2. sufficient voltage
    3. tape playback
    4. battery weakness

15. The condition of the batteries can best be determined by:

    1. the range the LEVEL meter pointer is in
    2. volume
    3. visual inspection of them
    4. distortion

16. If the pointer is outside the GOOD range:

    1. voltage is sufficient
    2. tape speed will improve
    3. batteries should be replaced
    4. the batteries are okay

Drills for wood or metal differ slightly in their shape, but their mode of action is almost identical, whether operated by hand or electrically powered. The wood drill (or bit) has a screw-like part at its center which drives into the wood and pulls the cutting edges after it. The cutting section approaches the wood with two sharply pointed knife edges that spear straight down into the wood and cut through the fibers in the cross grains so that they won't splinter. The cutting edges themselves, two knives which turn and slant into the wood, clean out all the wood between the spearing blades and carry it up along the twisted path on the body of the drill.

17. Which part of the bit enters the wood first?

    1. the cutting edges followed by the screw-like part
    2. the drive
    3. the screw-like part at the center
    4. the knife edge

18. Wood drills are powered:

    1. always by hand
    2. always by electricity
    3. sometimes by hand, sometimes electricity
    4. by the bit

19. Wood and metal drills are:

    1. different in shape and mode of action
    2. a little different in shape but much the same in action
    3. different in action but identical in shape
    4. exactly the same in shape and action

20. The passage is:

    1. about wood drills and metal drills equally
    2. mainly about metal drills
    3. is concerned with all types of drills
    4. is mainly about wood drills

Metal drills don't need the spearlike parts since metal has no grain and it will be cut smoothly just by the sharpened drill end. Usually the metal drill scrapes away the material at the very center of the hole and cuts it away out toward the periphery. It is driven into

Reading Comprehension EST (continued)

the metal by very high pressures and speeds in the drill press or even by hand-held electric drills. A hole larger than ¼ inch is usually made by first drilling a smaller hole and then finishing with a drill of the desired hole size.

21. Drilling a hole larger than ¼ inch in metal requires:
    1. spearlike parts
    2. the use of more than one drill bit
    3. a special drill press
    4. no grain

22. The spearlike parts found on wood drills are:
    1. unusual on metal drills
    2. characteristic of metal drills
    3. not a part of metal drills
    4. needed on all metal drills

23. Is there any reference made to wood drills in this passage?
    1. Yes, indirectly, in sentence number one
    2. No, there are no references at all
    3. Yes, there are references made throughout
    4. Yes, there is a reference made in the last sentence

24. One function of a drill besides simply drilling is:
    1. determining a hole size
    2. measuring
    3. waste removal while drilling
    4. avoiding high pressure

The important thing to remember when wielding a hammer is that the speed of the hammer head, not your muscle, is what gives maximum impact. It is more important to get a good stroke well placed than it is to get a roundhouse blow. Try to strike the nail with the hammer handle parallel to the work at impact. This means that your hand will be just a little farther away from the surface of the wall or cabinet than you would normally think is right. But by keeping the hammer face flat on the nail head on impact, the nail will drive true and straight rather than off to one side—and there's no wasted power. Also try to keep in mind that it is the head of the hammer doing the work, not the arm. Give it a good rotating swing from the elbow, not necessarily from the shoulder.

25. The best distance between the hand and the surface being worked on is:
    1. twice the length of the hammer
    2. less than most people think
    3. as near as possible
    4. more than most people think

26. To use a hammer with greatest efficiency
    1. a good roundhouse blow is important
    2. your hand should be close to the work surface
    3. speed and accuracy of stroke are important
    4. maximum muscle power is most important

27. Which of the following does *not* lead to wasted power when hammering?
    1. depending upon muscle strength
    2. keeping the hand very close to the object being hammered
    3. use of roundhouse blows
    4. use of a rotating swing from the elbow

Reading Comprehension EST (continued)

28. Which of the following does not contribute to maximum hammer impact?
    1. selection of the heaviest hammer available
    2. hammer head speed
    3. keeping the hammer parallel to the work at impact
    4. greater distance from the surface than you would normally think correct

A cell produces electricity by chemical means. That's the basic definition. The many available types have been divided into general kinds—primary and secondary. Those cells that are not rechargeable (flashlight D cells, for example) are primary cells, and those that may be recharged (automobile batteries) are secondary cells. The primary cell is disposable. It is so inexpensive that you can afford to throw it away and buy a new one when the old one has been drained. The secondary cell is rechargeable. When drained of current, it can be recharged by applying an external voltage supply to the terminals and reversing the chemical action that took place during discharge.

29. Cells that are not rechargeable are called:
    1. primary cells
    2. flashlight cells
    3. secondary cells
    4. automobile batteries

30. Secondary cells are:
    1. non-rechargeable and expensive
    2. expensive and rechargeable
    3. inexpensive and rechargeable
    4. inexpensive and non-rechargeable

31. This paragraph is concerned with a comparison between:
    1. D cells and other flashlight batteries
    2. expensive and inexpensive types of batteries
    3. primary and secondary cells
    4. automobile and flashlight batteries

32. Which of the following adjectives describes an automobile battery?
    1. secondary, non-rechargeable
    2. primary, rechargeable
    3. rechargeable, secondary
    4. non-rechargeable, primary

Part IV

Teacher Preparation

Turning Elementary Teachers Into ESL Specialists: A Cinderella Story?

Nancy R. Cott

Los Angeles Unified School District

Fraida Dubin

Univ. of Southern California

Surveys of graduate programs reveal that few elementary teachers receive certification or degrees in TESL/TEFL. It is more common for ESL teachers at this level to be recruited from regular classroom assignments. Elementary teachers, therefore, are a vital audience for in-service training in language pedagogy.

This paper discusses some sociolinguistic, sociopolitical, and attitudinal factors that appear to contribute to perpetuating a second-class status for elementary-level ESL teachers relative to their peers in academic or credential programs that are meant to train them, in the schools and society in which they work, and in professional organizations such as TESOL where their participation is often less than that of other special interest groups.

Drawing on their experiences as teachers and trainers, the authors reexamine some assumptions about course content and goals specific to this audience. They indicate the need to resolve the dilemma about whether to provide short-term training programs in response to teachers' expressed need for immediate, tangible, and practical help, or long-range education involving theoretical bases of language pedagogy designed to prepare trainees to function independently.

Since its inception in the mid-sixties, TESOL has been concerned with providing a common meeting ground for practitioners from every instructional level—elementary, secondary, college/university, and adult. By avoiding artificial divisions, this professional organization has emphasized its focus on the whole field of English language teaching. However, as the field has grown, so has awareness of the variability among learners as well as of teaching circumstances. English language teaching in 1979 is not uniformly the same wherever it is practiced. Thus, any consideration of the current state of teacher training must be stated in terms of a particular educational context or level.

Our interest here is in elementary-level public school teachers. How do they become ESL teachers? How are they prepared for the job? Our concern is that, like Cinderella, they appear to find themselves the step-sisters of their peers—in academic degree or credential programs that are meant to train them, in the schools and society in which they work, and in professional organizations such as TESOL. For example, compare the number of presentations at recent

TESOL Conventions which come directly from elementary ESL classroom experience with that from any other level. Out of almost 250 papers, panels, and workshops on the program of the 1978 TESOL Convention in Mexico City, there were only ten presentations directly concerned with the elementary level and most of these were in the context of bilingual education. Similarly, of the nineteen articles on teacher training listed in the *TESOL Quarterly Index* for Volumes I—X, none are specifically concerned with elementary level teachers.

What we see is a classic example of a self-perpetuating cycle: lack of solid professional preparation, lack of status and confidence as practitioners, lack of incentive or recognition for participation in academic activities or professional organizations, and lack of adequate representation, appeal, or opportunities for personal and professional growth at the institutional or organizational level.

The purpose of this paper is to explore the background of this situation. In so doing, we wish to share with other trainers, potential trainers, and elementary classroom teachers themselves (the Cinderellas), our experience as trainers of elementary-level ESL teachers. Our base for viewing this neglected audience in TESOL combines two decades of planning, organizing, and conducting courses for teachers, both as in-service classes and workshops within a school district, and as continuing education courses through a university. We also wish to validate our views with those of people doing similar work in other parts of the country.

What sets elementary teachers of ESL apart from their peers in other levels of instruction? Let us begin with the selection process. Most elementary ESL teachers come from the ranks of classroom teachers. Generally, they have had a regular teacher-training program organized around the traditional subjects of math, social studies, reading, and language arts. Language arts preparation is quite different from the content of TESL-based academic courses such as introduction to linguistics, or second language acquisition, in that it assumes that students are native speakers of the English language.

Prospective ESL teachers may be selected, they may volunteer for the job, or they may simply have a number of limited or non-English speaking students in their regular classroom. The basis for being singled out by school administrators is often because they speak the first language of the majority of the ESL student population, because they have an "ethnic" or immigrant background themselves, or because they express a special interest or concern. The motivation for volunteering may include a desire to get out of the drudgery of a regular assignment, a wish to be different, to have a special assignment or title, or a readiness to gain new skills and experiences. Rarely do elementary teachers enter ESL teaching at that level because of prior training or experience in TESL.

Occasionally, we encounter a "Pollyana" response which maintains that teaching ESL is like teaching any ordinary class: there is no need to learn a new job; regular methods and materials will do. We also sometimes find a patronizing attitude that *understanding* or *doing something for the poor children* is enough. Fortunately, many recognize the need for help and seek it. Whether desparately struggling for their own survival in a new and difficult situation, or altruistically seeking the best way to meet the special needs of ESL students, teachers who come for on-the-job training through in-service or extension classes are making a realistic response to a new challenge. It is our responsibility as trainers to recognize their needs and to help them make the transition. But unlike fairy-tales, there is no magic wand.

We must realize that this is a different population from those having certificates or graduate degrees in TESL. Our typical Cinderellas probably have a richer background in classroom management techniques and curriculum innovation (Taylor, 1978). They understand the complexities of the total elementary school instructional program. They understand the intricacies of organizing independent activities. They know how to get students to work in small groups. They are familiar with arranging learning centers and are concerned with making the classroom environment exciting and stimulating. They think in terms of realia and visuals. In fact, many graduates from TESL training/degree programs, well-versed as they are in language acquisition theory, could probably benefit from such expertise in these other areas.

This, then, is the usual background of the people who come for in-service training or continuing education. When we see them in the role of students, they manifest certain characteristics which add up to a composite picture of insecurity. Teachers often come as a group from their particular site, looking to one another for mutual support. Their expectations are set at picking up samples of everything. They freely express their need for help, but feel it should be in the form of something they can put into their hands rather than into their heads. They want quick recipes for what to do tomorrow, in the form of hands-on materials such as games and visuals. They often evaluate an in-service workshop in terms of how much they can take back and use themselves or put into a learning center for their students to use. We do not intend this to be a negative characterization. In fact, we see the give-me-one-of-everything reaction as a very normal and understandable response to the frustrations and realization of the enormity of their task. Our intent is to make trainers aware of the teachers' need for immediate, practical, tangible help in overcoming their sense of inadequacy.

In addition to these personal concerns, there are complex external factors contributing to elementary teachers' feelings of insecurity about being ESL teachers. There are any number of conflicting attitudes and opinions held by other teachers as well as the general public on the subject of language use and language learning in the public schools, some of which are enumerated below:

1) Why is ESL such a big issue? Haven't waves of immigrants throughout the history of this country learned English as a matter of course in the public schools?

2) What's wrong with the natural approach to language learning? Doesn't it happen by exposure, by osmosis? Why is there a need for directed instruction?

3) Isn't interfering with children's language damaging to their self-image? Why should we impose English on the child's natural growth and development?

4) How do we reconcile the recognition of students' right to their own language with the reality that they need English to achieve minimum competence for graduation and success in securing jobs?

5) Do the language needs of limited-English speakers differ from those of non-standard English speakers? Should they receive the same type of language instruction?

6) Does pulling out children for ESL class or to a language lab have a damaging effect by tracking, isolating, or labelling them? What are the alternatives, and what are the relative merits of different ways of grouping for ESL instruction?

Trainers who are prepared to address themselves to sociolinguistic questions such as these early in their program can usually establish a higher degree of mutual understanding and respect with an elementary-level audience. Trainers must also be prepared to deal with socio-political issues concerning the role of ESL in the context of bilingual education such as:

1) Advocacy or polarization of bilingual *vs.* ESL instruction and instructors, as if one negated or precluded the other.

2) An assumption on the part of some school boards and administrators that anyone who speaks English can automatically teach it. This ignores that fact that though they do read or use math in their daily lives, teachers have had to undergo training and qualify for credentials in order to teach these subjects.

3) Bilingual teacher certification and programmatic requirements specify fluency and literacy in the students' language and knowledge of the culture associated with that language but tend to omit any standards for competence in English, or knowledge of how language is learned or taught.

4) Similarly, and possibly consequently, recruitment and in-service efforts have developed the same emphases and thus virtually ignore the whole population of monolingual or other teachers who could and should be identified or trained to teach ESL.

5) Staff development activities are often devoted to explanations of the intricacies of legal requirements and criteria for compliance with state and federal mandates but seldom spell out how these goals are to be carried out in the classroom. At best, they are prescriptive or proscriptive, but give the teacher little choice or direction; individual practitioners have to figure out how to do it themselves.

6) There is ambiguity and ambivalence about where ESL belongs—whether it is considered a part of the bilingual program or a legitimate part of the curriculum with a content, scope and sequence of its own. ESL instruction competes for dollars, staff, materials, and space with primary language instruction in bilingual programs on the one hand, and conventional instruction in basic skills on the other.

7) Teachers feel frustration when they have insufficient time on a regular daily basis, or for a duration of longer than one school year, to give students enough specific ESL instruction to see concrete results. The absence of demonstrable positive outcomes leaves teachers with feelings of dissatisfaction about their jobs, and concerns about keeping them.

Thus, we see that ESL teachers in the context of bilingual education may feel resentful or cynical. They sense that the scene is charged with as many emotional and political questions as pedagogical ones. They see themselves in a second-class status in relation to other program instructors, licensing and certification, instructional priorities, and job security. Given these and other unresolved issues, is there any wonder that new and prospective elementary-level ESL teachers aproach their tasks tentatively?

And given this complex array of factors affecting instructional concerns, is it any wonder that training programs originating in academia and operating only on a theoretical or abstract level usually fail to attract or hold an audience of elementary-level ESL teachers? An understanding of what is appropriate for this audience is the trainers' real dilemma and challenge. Experienced trainers recognize that there are no simple, clear-cut answers to the many questions in the field, yet they must make decisions about their own roles and goals.

The trainers' dilemma over whether to start with theory or application was faced some years ago when Peace Corps TESL/TEFL programs were in their prime. Campbell (1967) discussed the distinction between long-and short-range programs, and between "education" and "training". He pointed out that if trainees understand basic principles of language and language pedagogy, then they are able to answer their own questions. However, this kind of understanding takes considerable time to acquire; trainees still need something to use tomorrow.

How do we reconcile the needs for long-range education vs. short-term training in working with elementary-level teachers? There are two main areas of consideration: assumptions about course content, and attitudinal factors. Many conscientious trainers believe that their ultimate goal should be to get trainees to the point of answering their own questions and making their own decisions. But they automatically assume that the first step, or lesson one, should be a lecture on the nature of language, followed by assigned reading and further lectures designed to provide in-depth background in the fields of second language acquisition, sociolinguistics, grammatical theory, and on issues of language pedagogy, cross-cultural awareness, and curriculum planning. Even when such courses begin with theory and promise to get to application later, the result can be alienation, hostility, or no return audience next class or next term.

In practically every group of trainees, there are some individuals who are stimulated by an abstract approach and who respond positively rather than being bored or antagonized. However, back in the classroom, they run the risk of even greater disappointment if their increased awareness is not tied to practical concerns. If, on the other hand, trainers try to be responsive to teachers' needs by providing them only with survival techniques and materials for immediate classroom use, are they being responsible or responsible for simply perpetuating teachers' dependence on hand-outs?

Trainers should examine themselves for latent or implicit signs of condescension when addressing teachers "down" at the elementary level. Often unconsciously, they express the attitude that the higher the instructional level, the higher the intellectual capabilities of the audience. Trainers who talk over the heads of, or down to, elementary teachers show some striking parallels to inexperienced teachers of ESL who often speak too distinctly, loudly, or slowly to their students.

Trainers should be honest about their own background and experience. If they have had little or no exposure to the realities of the trainees' jobs, in terms of locations, organization, time constraints, or programmatic requirements, they are at a disadvantage. However, if trainers are working without first-hand knowledge of the trainees' circumstances, it is best to be straightforward about the situation and ask the trainees to supply necessary background information. Sensitive trainers should come away from this experience feeling that they have learned as much from the participants as the other way around.

Trainers need to know if they are facing a captive or a self-selected audience. They should also differentiate between an audience composed of novices as opposed to more experienced people. But no matter how diverse the audience, trainers can usually count on there being some who want how-to-do-it formulas, while others are looking for more general background and basic principles. Both groups should be provided for. The challenge for trainers is to come up with things that trainees can learn from themselves as well as use with their students.

Trainers need to come out of the ivory tower. Successful programs draw on practical examples which relate to the trainees' own experiences. Here again, a parallel can be drawn with good ESL instruction that goes beyond frontal teaching and empty drills to involve students in realistic and meaningful contexts and situations. In this connection too, trainers need to decide if and to what extent they should function as a sounding-board for teachers' anxieties and socio-political concerns.

In sum, the best kind of training program should provide a model itself. "The medium is the message" aptly applies to teacher training. The processes through which training takes place can be used as examples of how to teach language to elementary school children.

A vital goal in such a training program for elementary-level teachers is to help eradicate the Cinderella complex. The teacher who possesses confidence that comes from appropriate training and successful experiences is ready to enter into professional activities. We hope that at next year's TESOL Convention, we will see more presentations directed to this audience, and more elementary teachers trying on "Cinderella's slipper" and addressing themselves to issues of their neglected level.

ESL Teacher Training: The Need for a Complete Curriculum

Jacqueline K. Neufeld
University of Cincinnati

Marion R. Webb
Houston Baptist University

The growing demand for ESL teachers, caused both by legislatively mandated classes for United States residents of limited English-speaking ability and by the influx of foreign students into the country, has resulted in the addition of ESL components to teacher training curricula in many colleges and universities. Often these ESL components function under the tutelage of already existing curricula in Foreign Language or in English, where ESL becomes something of an academic stepchild, with similarities stressed and differences ignored. Yet ESL students in monolingual, multicultural programs are quite different from students of English as a first language, foreign language students, or ESL students in bilingual classes. Nor are the monolingual ESL programs themselves the same as those in English as a first language, foreign language, or bilingual education. The dissimilarities include, among others, amount and sources of funding, student motivation, geographic setting, ages of students, levels and orientation of languages taught, and cultural emphases. Because of the growing demand for ESL teachers, and in view of the fundamental differences between ESL programs and other language programs, this paper argues for the development of ESL teacher training programs unaffiliated with English, or foreign language, or bilingual education, and for state certification of ESL teachers, in order that ESL teacher training may be expanded into the complete curriculum that it deserves to be.

Programs for the education of all language teachers clearly share many areas of preparation. The overlapping of subject matter, materials, and aims has long made the fields of teaching English to Speakers of Other Languages, teaching the various foreign languages, and teaching in bilingual programs interdependent from the viewpoint of research and cooperation. Much of the recent literature cites common ground and the sharing of common concerns among our disciplines (Alatis 1976b; Kunkle 1977; Finocchiaro and Ekstrand 1977). While we endorse all possible cooperation among our interrelated fields, we feel that there is a growing need for English as a Second Language (ESL)[1] teacher preparation curricula that are separate and independent from those for bilingual education teachers, those for foreign language teachers, or those for teachers of English as a first language.

---

[1] We are aware of the technical differences implied in the designations English as a Second Language (ESL), English as a Foreign Language (EFL), and the generic English for Speakers of Other Languages (ESOL). In this paper we have chosen to use ESL generically because it is the more popular designation with most educators, and because it is appropriate with respect to order of language acquisition.

Our premise is that students of ESL who are not enrolled in bilingual education programs are different from those who are and different also from typical American foreign language students. Further, these ESL students are the fastest growing body of language students in the U. S. today. Our conclusion is, therefore, that these students deserve to be taught by language teachers thoroughly prepared in programs that, while sharing similar aspects with bilingual education, foreign language education, and English education, are, for the most part, independent of these, just as these curricula are independent of each other. It follows, then, that such teacher candidates deserve in turn to be prepared in programs directed and taught by those thoroughly versed in all aspects of TESOL.

Three main groups of ESL students not enrolled in bilingual programs can be identified: foreign students who study ESL in our colleges and universities; ESL students in Adult Basic Education programs, and elementary and secondary students enrolled in Multicultural or ESL programs. A quick look at the numbers involved will give an estimation of the need for professionally trained ESL teachers.

Concerning the first group, the annual survey of the Institute of International Education reports that during the 1977-78 academic year there were 235,509 foreign students on U.S. campuses, an increase of more than twice the enrollment of ten years ago (*IIE Reports* 1979:5), and an increase of 15% in only one year (*Time* 1979:63). While not all of these students are enrolled in ESL classes, it is safe to assume that the majority have been enrolled in at least one ESL class since their arrival in the United States, and many have been studying nothing but ESL for a year or longer. The number of foreign students studying in the U.S. in 1977-78 was almost equivalent to the number of students in the U.S. who studied university level French two years before that. And if only one-third of the foreign student population was enrolled in ESL at any one time, there would still be as many as studied German in all the colleges and universities in the U.S. two years ago (based on enrollment data reported in Brod 1976:169).

The second group, ESL students in Adult Basic Education, is potentially the largest group of all. According to estimates based on the Survey of Income and Education conducted by the National Center for Education Statistics of the Department of Health, Education and Welfare, approximately 2.4 million persons in the U.S. speak no English at all, and of these 90%, or 2.16 million, are adults nineteen years of age and older (Waggoner 1978:252). Instruction to equip adults with basic literary skills is provided as a part of the Adult Basic Education Act funding; English for adults who are non-English speaking is interpreted as a basic literacy skill.

Finally, the third group is comprised of those students in elementary and secondary schools from non-English speaking backgrounds who are not enrolled in bilingual/bicultural programs. The survey referred to above identified more than ten million children between the ages of six and eighteen who are of non-English language backgrounds (Waggoner 1978), although not necessarily non-English speaking. Clearly many of these students will be enrolled in multilingual/multicultural programs, or some type of ESL program. The 1975 *Task Force Findings Specifying Remedies Available for Eliminating Past Education Practices Ruled Unlawful under Lau v. Nichols* specify that "a district does have an obligation to serve any students whose primary or home language is other than English" (p. 4).[2] Even discounting from this number of 10 million those children who are in bilingual/bicultural programs, and those who have become English-speaking, the number studying in multilingual/multicultural programs or in ESL programs according to the specified guidelines ought to be large.

Further endorsing the idea that the ESL population in the U.S. is large, Kunkle (1977:213-214) identified 4,913,000 non-native Americans and residents who have serious difficulties with the English language. In his article he states that "a serious commentary is that of this group of almost 5 million people with difficulty in English, approximately 4 million have not been enrolled in school at any time since September, 1974 (Kunkle 1977:214). If this is to be recognized as a problem, as undoubtedly it is, then we must ask ourselves, as professionals, what is the source of this problem?

A major contributing factor not only to this problem but also to ones dealing with funding and recognition of ESL on campuses and in state supervisory offices is the program's uncertain status. Fanselow and Light (1977) present the historical perspective of why this has been so. They state that after World War II

> in spite of the fact that the teaching of English to non-native speakers of English had been expanding greatly . . . , it was difficult to know during those years what the needs of the field were and what the qualifications of the teachers were or should be. One of the reasons for this no doubt was the widespread attitude that speakers of a language can teach that language to non-native speakers by virtue of the fact that they speak the language! Another reason, perhaps a result of the first, was that state education agencies usually had no offices to handle this area. Even on campuses where English was taught to foreign students, the status of the programs and the teachers in them was often not clearly defined. (p. 3)

We have not outgrown this heritage. We still lack a national concensus as to which departments should house programs for ESL students, which departments should house the programs that prepare the teachers of these students, and which state and local offices should supervise the programs that give direction to the teachers in them. In fact, the present situation can be described as little short of chaotic!

---

[2] In regard to the role of English as a Second Language in elementary bilingual programs, Nakano (1977:219-234) notes that "the single most inflammatory and injudicious statement made in the OCR Remedies is that:

> Since an ESL program does not consider the affective nor cognitive development of the students in this category and time and maturation variables are different here than for students at secondary levels, an ESL program is not appropriate (p. 232)."

This statement is unfortunate, since the 1976 TESOL Resolutions on Certification specifically call for a cultural component in TESL certification (Fanselow and Light 1977:250-251). The task force findings, however, specify some type of "plan" (bilingual/bicultural, multilingual/multicultural, transitional bilingual education, or high intensive level training—the latter at the secondary level only) "when the district has 20 or more students of the same language group identified as having a primary home language other than English" (p. 4).

Across the nation we find ourselves as a profession subsumed under such faculties as Foreign Language, Bilingual Education, and English, none of which can ever prove entirely satisfactory. Often ESL teacher preparation does not have a complete curriculum, but functions under the tutelage of already existing curricula in foreign language or in English where it becomes something of an academic stepchild, with similarities emphasized and differences ignored. At state and local public schools levels, ESL supervision and program development may be housed under English, or under foreign languages, or under language arts, or under bilingual education, or responsibilities divided among these various offices, or all of these combined into one office. In a survey of state foreign language supervisors, we asked where ESL supervision at the state level was housed. Of the twenty-three responses not one responded that ESL was housed in its own separate department or office. The May 30, 1978 issue of the *Chronicle of Higher Education* addressed this same problem.

> The uncertain status of EFL has led to debates over who should be teaching it. On some campuses, the courses are found in foreign language departments; on others, in regular English department. Still others view the courses as college preparatory. (p. 7)

The uncertainty about who should be teaching ESL is caused, to a great extent, because people outside the field do not have a clear definition of the scope, practices, and theoretical bases of teaching ESL. Because of the unfamiliarity with what we do, as well as a lingering hint of mystery, campuses and supervisory boards are often very unclear as to what department or even to what discipline we belong.

As professionals, we need to define the borders of our profession in such a way that no uncertainty will remain as to who should be teaching non-native English speakers. Naturally there is a reason why so many ESL programs are housed in these different departments across our country. Commonalities do exist; however, because there are great differences separating the fields, we propose a rationale for separating ESL programs into separate entities taught and staffed specifically by people with expertise in ESL. The impressive numbers cited as well as the lack of certainty about who should be teaching ESL students and what our status is support the need for sound ESL teacher training curricula. Strengthening the rationale that such curricula should be separate from those designed to train bilingual teachers and foreign language teachers is the fact that ESL students themselves are quite different from other language students. Some of these differences are:

1) *Geographic setting.* As basic as this distinction seems, it is surely the most fundamental difference between ESL students in the U.S. and other language students. Adult ESL students generally are at sea in a foreign culture which at the very least demands of them survival skills in English. In contrast, English as a first language students have their support systems—family, friends, culture, as well as the linguistic environment—all intact. American students in

foreign language programs also have these. Generally, their language study is not for immediate use but rather serves the goals of a humanistic education or some future experience. (Exceptions may be geographic areas where the study of Spanish, or Creole French, or Haitian French has immediate and utilitarian purpose.) Even students in bilingual programs have both support systems and linguistic-cultural systems that they can fall back on. In addition, students in bilingual programs will have teachers and many peers who are knowledgeable about and sympathetic to their cultures. Because ESL students probably are enrolled in classes composed of many nationalities, they may have no class member nor a teacher who can speak or understand their languages, and they may have no one who understands their cultural background.

ESL students also differ from typical foreign language students in their opportunities for the acquisition of language. The acquisition of English by ESL students is an inevitable result of the daily environment, whether they are at school, at work, watching television, shopping, or participating in leisure-time activities. They thus live in what Krashen (1977) calls an *acquisition-rich* environment, as contrasted to the *acquisition-poor* environment of typical foreign language students. As Krashen notes, however,

> the adult second language learner has two means for internalizing rules of the target language. One is language acquisition, which is very similar to the creative construction process utilized in both child L1 and L2 acquisition. . . . The other process available to many adults is language learning, which results in conscious representation of pedagogical rules (p. 145).

Given these two processes, ESL students have the advantage over foreign language students, because while "acquisition may be slow . . . it is, in the long run, much more useful when language is used for the purpose of communication" (p. 156).

2) *Age.* We have discussed language learning versus language acquisition in terms of the adult language learner. When the number of foreign students is added to the number of persons in the U.S. who are non-English speaking, the total is more than 2,635,000 persons who are potential ESL students, less than ten percent of whom are nineteen years of age or under (based on the data in Waggoner 1978). This does not include, of course, the more than ten million children between the ages of six and eighteen who are from non-English language backgrounds. Many will be enrolled in bilingual programs, and many do speak English. Thus, most ESL students, in contrast to the majority of students in bilingual programs and most students in foreign language programs, are largely adult learners.

3) *Motivation.* Because ESL students are in the United States, and because they are for the most part older students, they are generally motivated to learn English as a means of achieving immediate ends—whether these are to gain a college education, to complete a graduate degree, or, as new immigrants or businessmen, to survive in the English-speaking world. This motivation is what Gardner and Lambert (1972) call *instrumental* motivation, rather than the *integrative* motivation typical of some of the best language learners in bilingual

education programs. In any case, ESL students are in this respect different from English as a first language students, who take for granted the very automaticity that ESL students are trying to develop. They are different as well from American foreign language students, who typically conceive of foreign language as a course in the curriculum, rather than a means to achieving necessary survival skills.

4) *Cultural Emphases.* For the most part, the ESL class, unlike the bilingual education class or the foreign language class, brings together not just two cultures, but a rich variety of cultures. In this multicultural environment ESL students share their cultures with others while at the same time learning to appreciate the universality and the shared human values of all cultures. Yet the common bond of language and culture shared among all class members is the English language and the American culture, both of which are embodied in the teacher.

5) *Levels and Orientation of Language.* ESL students will generally study English longer than foreign language students will study foreign languages. Therefore, programs for ESL students will include more levels of study and more skill development in all areas of English. There will usually be less attention to literature in programs for ESL students than in foreign language programs. Because of the mixture of nationalities in the classroom, the orientation of classes and of materials will usually be in English only, rather than English and the native language. Additionally, the progress of motivated ESL students can be more rapid than that of foreign language students because ESL students are learning and acquiring the language.

6) *Sources of funding.* ESL students at the elementary or secondary level can attend tax-supported tuition-free schools (unless they happen to be undocumented aliens). ABE-ESL students also have no tuition to pay. But aside from a small percentage of exchange students, other adult ESL students and college ESL students may pay their own way or their countries or corporations must pay their tuition. While American college students have a number of scholarship sources, ranging from local to federal funding programs, and from private scholarships to public grants, ESL students must bring ample funding along with their visas, or, at the graduate level, have sufficient command of English to secure teaching assistantships or other means of support.

7) *Research.* Because ESL learners in general are in an environment in which language acquisition may be more important than language learning, much of the advancement of second language acquisition research theories has resulted from studies of ESL learners, both children and adults (Bailey, Madden and Krashen 1974; Dulay and Burt 1974; Larsen-Freeman 1975). On the other hand, while many studies of learner attitudes have enlisted the participation of both foreign language students and of those in bilingual programs (Lett 1977; Cooke 1978; Gardner and Lambert 1972), there has been little attitudinal research at beginning levels of ESL. While the culture shock that foreign students suffer has been observed and reported (Bedingfield 1978), we have little empirical knowledge to tell us how foreign students feel about learning English at beginning levels or how they feel about American culture. In most programs,

instruments designed to tap these attitudes would need to be translated into scores of different languages, and reliability of the translation into each language established, a time-consuming and expensive task.

We have described several of the ways in which ESL students in monolingual programs are different from ESL students in bilingual programs, from foreign language students and from English as a first language students. These, along with a few other obvious differences, are outlined in table form in the appendix. (English as a first language is omitted from this table because, for the most part, the differences between it and ESL are much greater and more obvious than those between ESL and foreign language or bilingual programs.) In view of the differences discussed between ESL students and those in other language programs, what are the implications for the training of those who will teach ESL students?

In the first place, unlike the typical foreign language teacher, the ESL teacher will not be the sole determiner of the foreign language content that the student will be exposed to because ESL students will be learning language in the classroom as well as acquiring it in the process of daily life in the U.S. Therefore, the ESL teacher will need to know in what ways language acquisition outside the classroom can best be utilized to further the learning process in the classroom.

Second, because the majority of ESL students are older students, the ESL teacher will need to be aware of teaching methods suitable to that age group. Many foreign students and many adult ESL students will be highly motivated to learn English, but at the same time may have false hopes about the length of time it will take to become fluent in English. The ESL teacher will therefore need great patience and understanding. Thus, the elitist attitudes once prevalent among many foreign language teachers clearly have no place in the ESL classroom.

Third, because of the multicultural nature of most ESL classes, the ESL teacher will need to be knowledgeable about many cultures and evince respect for students from all of them. In the bilingual classroom, the teacher is expected to be bilingual and bicultural. In the ESL classroom, at least in the typical one where students are of many language backgrounds, the teacher will not often be multilingual, but she or he can become an appreciator of many cultures. Moreover, because the ESL teacher cannot speak all the languages represented in the class, she or he has an even greater need to have a thorough knowledge of the way all languages work—that is, a thorough grounding in descriptive linguistics, applied linguistics and sociolinguistics. And because English will generally be the only language of the ESL classroom, the ESL teacher will need a thorough knowledge of all aspects of the English language—that is, syntax, morphology, and semantics.

Due to the factors just discussed, it is logical to assume that if ESL students have unique characteristics, and if ESL teachers need unique training, then there should also be unique qualifications for future ESL teachers. Therefore, we should expect the curriculum designed to prepare future ESL teachers to be different from that designed to prepare future foreign language teachers or future bilingual teachers. This is not, however, the case. Although the *Guidelines for the Certification and Preparation of Teaching English to Speakers of Other Languages in the United States* were ratified by TESOL in 1975, there seems to be, as yet, little concensus as to what a core ESL program should entail. The rapid growth in the number of students in ESL in recent years has created tremendous demands for ESL teachers. According to the preliminary data of the 1976-77 survey of the National Center for Education Statistics (Waggoner 1978:257), there were more than 100,000 teachers of ESL. Yet only three in ten had taken even one course in ESL! Only 125 institutions of higher learning were tentatively identified as offering curricula that require courses predominantly in teaching ESL.

As a result of these figures and because many foreign language teachers are assigned to teach ESL, we asked the state foreign language supervisors that we surveyed to identify areas of ESL teacher preparation that foreign language teachers would most need to study if they were to be employed as ESL teachers. Then we asked our audience of ESL educators and teachers who attended our session at the annual TESOL meeting in March, 1979, to respond to the same question. The ten areas of preparation that received the most votes by each group are listed in their order of choice in Table 1. As is readily apparent, while similar courses appear in both columns, nevertheless the focus by those specifically teaching ESL or preparing teachers of ESL is different from the focus of those tangential to the field. This difference no doubt reflects the fact that priorities for ESL teacher preparation have not been agreed upon.

TABLE 1

Areas of ESL most important for a foreign language teacher to study

Choices of state FL supervisors ($N = 23$)	Choices of ESL educators/teachers ($N = 31$)
ESL methods	ESL methods
Phonology and morphology	Reading
Testing and evaluation	Psychology of the L_2 learner
American culture	Testing and evaluation
Syntax	Syntax
Psychology of the L_2 learner	Theories of language development
Reading	Review of current literature and research
Theories of language development	Phonology and morphology
Composition	Sociolinguistics
Psycholinguistics	American culture

Moreover, in too many colleges and universities, only a course or two is offered to prepare future ESL teachers. A complete curriculum is lacking. It is obvious, however, that teacher training institutions will establish complete curricula to meet state requirements if states mandate certification or endorsement for ESL teachers. Blatchford (1977, 1979a) has extensively surveyed the college and university programs for training ESL teachers as well as the present status of state requirements for certification or endorsement in ESL. In his most recent survey, Blatchford (1979a, 1979b) found that only eleven states currently have certification or endorsement for ESL teachers. In order to determine if there was a nationwide trend toward implementing some type of ESL certification at the state level, we asked the 41 state foreign language supervisors identified by ACTFL (1978) if they foresaw the addition of ESL certification in the near future. It was our hypothesis that most states would not have certification requirements specifically for ESL nor would they plan to add such requirements. Out of 23 returns, 16 responded that their states did not have now, nor would they have in the near future, ESL certification requirements. Our hypothesis that there is no trend toward implementation of state certification requirements for ESL was confirmed at the .05 level by a chi-squared analysis. If one considers the universal requirements for a complete curriculum of preparation and for certification of English as a first language teachers and for foreign language teachers, then it is obvious that ESL students are being discriminated against. The *Sample Rationale for ESOL Certification* developed by the Task Force of the New York State English for Speakers of Other Languages/Bilingual Education Association states that:

> Non-native speakers of English of all ages who have classes in ESOL are denied the right of having teachers certified in ESOL teach them ESOL. The students in chemistry class have teachers certified in chemistry, the students in French class have teachers certified in French, but the students of ESOL classes do not have teachers certified in ESOL because there is no certification in ESOL. . . . (Fanselow and Light 1977:257)

If the inequity to ESL students resulting from too few states requiring ESL certification can be remedied, then teacher training institutions will be in a position to offer complete curricula in ESL.

Given the uncertain status of ESL at present and the growing number of ESL students, and given that there are many differences between these students and other language students, can we then not argue for complete preparation and certification of ESL teachers? Must we not demand ESL preparation and certification as distinct from preparation and certification of foreign language teachers, bilingual education teachers, and teachers of English as a first language? Indeed, we must and do. But not only should we demand certification of our teachers, but we must, as a professional organization, insist upon setting the curriculum for our future teachers. As a professional organization we should assume responsibility for instituting a standard body of knowledge that all teachers trained and certified in ESL would share. Increasingly, teacher organizations, like other professional associations, are asking to control certification of those entering the profession. We as specialists, also need to accept this responsibility so that we may ensure quality ESL instruction for our students. We, after all, are the ones most aware of the ways in which the ESL learner is different from the learner in bilingual education programs, the foreign language learner, or the learner in English as a first language classes. We must, therefore, not only insist that ESL teacher certification or endorsement be required, but we must take an active hand in prescribing the curriculum necessary for future ESL teachers. Without the active involvement of the TESOL profession in delineating its boundaries, quality ESL instruction as well as its status will always remain uncertain.

APPENDIX

OUTLINE OF GENERAL AREAS OF DIFFERENCE AMONG PROGRAMS IN THE U.S.

Jacqueline K. Neufeld
University of Cincinnati

Marion R. Webb
Houston Baptist University

	ESL (MONOLINGUAL)	FOREIGN LANGUAGE	BILINGUAL
FUNDING	Funding due to bilingual acts and as a result of *Lau vs. Nichols*. Funding generally excludes non-residents and aliens who compose the majority of ESL students in American universities.	Heavy funding in the '60's from NDEA grants as a result of post-Sputnik need to improve math, sciences, etc. as a direct result of the Cold War.	Heavy funding from Bilingual Acts as a result of various court decisions. Bilingual students generally U.S. residents and therefore qualify for funding.
MOTIVATION OF STUDENTS	Ultimately pragmatic for purposes of business, academics and professional, whether instrumental or integrative. Monetary opportunities often part of motivation.	Aesthetic; liberal arts requirement; travel. (Spanish exception to this in some areas.)	Assimilation; peer pressure to speak like other children of same age.
GEOGRAPHIC	Natives of L_1 study L_2 in L_2 environment.	Natives of L_1 study L_2 in L_1 environment. (Some regional exceptions.)	Natives of L_1 study L_1 and L_2 in L_2 environment.
AGE OF STUDENTS	Wide range of ages, especially in elementary and university levels.	Mostly high school and university enrollment.	Mostly elementary age children.
LEVELS	Especially in university programs, more students in advanced levels.	In both high school and college most students in first and second year level.	Basic concept formation; development of academic basics.
MATERIALS	All levels; at advanced college level, great concern with perfecting reading and writing skills.	All levels, but majority of students at beginning levels. At advanced levels, great concern with literature.	All levels; emphasis on developing reading and writing skills.
CULTURAL EMPHASIS	Multicultural; emphasis on understanding culture so as to effect better communications competence.	Appreciation of L_2 culture.	Maintenance of cultural identity of native language while learning to appreciate culture of L_2.

LANGUAGE BACKGROUND OF TEACHERS	Majority of teachers L_1 native speakers teaching L_2.	Majority of teachers L_1 native speakers teaching L_2.	Majority of teachers bilingual, teaching L_1 and L_2.
PROFESSIONAL BACKGROUND OF TEACHERS	Majors in: 1) Foreign languages 2) Elementary/secondary education with emphasis on English 3) Few majors in ESL	Most college majors in foreign languages.	Majors in: 1) Elementary education with emphasis English 2) Elementary education with emphasis bilingual 3) Foreign language
LANGUAGE ACQUISITION THEORIES	L_2 language acquisition similar to L_1 acquisition.	L_2 language acquisition not similar to L_1 acquisition. (cf. especially Brooks (1966).	L_2 acquisition similar to L_1 acquisition; emphasis on natural sequences in child language development.
RESEARCH	Emphasis on study of language acquisition in both children and adults; little use of L_1 survey research to investigate attitudes toward L_2.	Emphasis on study of language learning, largely in high school and college learners; use of surveys in L_1 to investigate attitudes toward L_2.	Emphasis on child language acquisition; studies of effects of bilingual education on basic achievement.
PEDAGOGICAL APPROACH	Integrative approach; stresses survival in American English-speaking communities and in American school and college system.	Assumes a homogeneity of student populace; not geared toward teaching academic survival needs.	Integrative approach; maintenance of L_1 while developing fluency in L_2.
TESTING	Mostly integrative approach; not concerned with piecemeal knowledge.	Mostly discrete point rather than integrative in large majority of first and second year classes.	Language dominance testing; basic skills in L_1 and L_2.
WHERE HOUSED	Subsumed under other departments such as foreign language, English.	Separate department or facility.	Separate department or facility or housed with language arts, foreign language, English or K-12 education.

A Model For Training The Non-Traditional TESL Student

Mark A. Clarke
University of Colorado at Denver

Bernard Seward
Bridge International School

With the recent influx of non-English speakers to many U.S. urban areas—primarily Indochinese refugees and Spanish-speaking immigrants—many non-ESL teachers find that they need TESL skills to deal with their pupils. As a result, many ESL teacher-training programs attract non-traditional TESL students such as bilingual education teachers, paraprofessionals, adult basic education volunteers, and public school teachers with non-English-speaking pupils in their classes. The challenge to teacher trainers is to prepare a compact, one-term course to meet the needs of such a diverse population.

Part One of the paper reviews the recent literature on teacher training in ESL and provides a summary of the main issues as they are revealed in this literature. A bibliography is included.

Part Two describes a one-semester "non-traditional" TESL methodology course developed at the University of Colorado at Denver to meet the needs of a diverse non-traditional TESL student population. It is argued that teacher trainers need to focus on specific techniques, and to address theoretical issues only when they arise in discussions of those techniques. In presenting a technique, trainers should make explicit 1) the ESL proficiency level, 2) the skills and 3) the type of pupil for whom the technique is appropriate. A three-component model is presented for implementing this approach. A course syllabus is outlined and sample classroom procedures are described.

1. Introduction

If our experience at the University of Colorado at Denver is indicative of a trend throughout the United States, teacher training programs will increasingly have to offer courses to meet the needs of non-degree students seeking skills in teaching ESL. Many of these students are teachers untrained in TEFL/TESL methodology and linguistics who have recently discovered their classes filled with non-English speaking pupils. Faced with this situation these teachers have two alternatives: (1) ignore the needs of the pupils or (2) seek assistance in acquiring the knowledge and skills to work with non-English speakers. Much to their credit, these teachers in ever-increasing numbers are choosing the latter alternative.

The challenge to teacher trainers is to prepare a compact, one-term course to meet the needs of these non-traditional TESL students. Part One of this paper reviews the recent literature on teacher training. Part Two presents a model for a one-term TESL course in which theoretical issues and practical classroom techniques are integrated.

2. Review of literature[1]

There is considerable agreement in the literature on the question of what constitutes the ideal ESL teacher. Early articles by Brooks (1966), Cloze (1966) and Marqwardt (1969) emphasize the interdisciplinary nature of the training of second language teachers, a view which is echoed later by Strevens (1974) and Alatis (1974). Alatis' article is representative of these points of view. He offers the acronym LAPSE as a framework for teacher preparation: *L*inguistics, *A*nthropology, *P*sychology, *S*ociolinguistics, *E*nglish/*E*ducation. Guidelines for the preparation of language teachers have been developed by the Modern Language Association (1966), the National Council of Teachers of English for English Teachers (1968) and TESOL (1975). The guidelines vary in a number of details, but all attempt to list the personal and professional qualities of the ideal teacher. In general, they advocate entrance criteria which assure native or near-native proficiency in the target language and personal qualities which would allow the individual to understand and take advantage of cultural differences in the classroom. In addition, they present outlines of course offerings which cover a wide range of academic areas such as those contained in Alatis' LAPSE approach. They also recommend some sort of practical experience in language teaching as part of the trainee's program.

While educators seem to agree on the qualities of the ideal teacher, there is less agreement on the most effective procedure for training teachers. Articles which focus on the process of teacher training can be divided roughly into two categories: 1) those which are Methods oriented and 2) those which are Technique oriented.[2]

The Methods-level articles grapple with the issues of program development and general orientation of training programs. Strevens (1974) highlights the characteristics of a good teacher and sketches the outline of a training program. Alatis (1974) goes one step further and reports on the evaluation of an

---

[1] This review is by no means exhaustive; we attempt to highlight recent work which might offer specific programmatic suggestions for the teacher trainer. An excellent review of the issues and practice in (and a thorough bibliography of) teacher training in ESOL, bilingual and foreign language teaching is contained in Fanselow and Light (1977). Fanselow and Light make an important point in the introduction to the book: "In reviewing the literature on teacher preparation in the journals related to language teaching, not only is there a dearth of articles of a research nature, but in fact, there are very few selections that treat teacher preparation at all." (1977, 9). It might be added that the majority of the articles which do address teacher preparation are of a theoretical or issues nature; there are very few which provide specific suggestions for teacher trainers.

[2] "Method" and "Technique" are used loosely here in the sense developed by Anthony (1963). In this sense, the guidelines mentioned earlier are on the Approach level because they embody seldom-questioned truths or axioms of teacher preparation. The Method level articles focus on general procedures and content while the Technique level articles focus on specific strategies and activities. It is assumed, of course, that in a particular program, there would be complete congruity among the three levels; that is, the techniques used would proceed from a methodological point of view which is firmly based on a well-developed approach to the training of language teachers.

M.A. program at Georgetown University. Lee (1974) and Taylor (1978) argue for a more practical focus in teacher-training classes while Jarvis (1972), Combs (1972), Blatchford (1976) and Dowling and Shepard (1976) emphasize the need to focus on the subjective, humanistic side of teacher training so that teachers can develop the ability to make accurate assessment of their students' problems. A substantial body of literature on the Method level is devoted to the pros and cons of competency and performance-based teacher education (CBTE and PBTE). Growing directly or indirectly out of researchers' attempts to identify the classroom behaviors of good teachers (see, for example, Fanselow 1977, Lee 1972, Moskowitz 1976) the movement for CBTE/PBTE is founded on the belief that basic competencies can be identified, isolated, and then imparted to prospective teachers in a teacher training program. Supporters of one form or another of CBTE/PBTE (for example, Fanselow 1977, Lee 1972, Schmieder, et al. 1977) point to the need for objective criteria for judging the acceptability of the training of future teachers. Critics (for example, Blatchford 1977, Resolutions 1977) point out that there is a great deal of work to be done before we can confidently isolate specific behaviors which characterize good teachers. Indeed, in a review of the research literature Cyphert (1972) comes to the conclusion that there is little available empirical data which illuminate the relationship between teacher behaviors and student progress. Although CBTE/PBTE-related articles usually offer quite detailed recommendations, they are classified as Method-level because they rarely provide specific suggestions for teacher training techniques.

Articles on the Technique level cover a wide range of issues and techniques in teacher training. Orwen (1971) offers a blueprint for the Methods course, providing a framework for team teaching by trainers and peer teaching by trainees. McArdle's (1971) article cuts across the Method/Technique classification we have established but is included here because he offers a detailed framework for evaluating the knowledge and skills of trainees. A number of articles suggest refinements in microteaching, a technique in which trainees teach a short (five to fifteen minutes) lesson to a group of ESL students, usually focusing on a specific language problem. Politzer (1966) is widely recognized for his use of micro-teaching techniques. Modifications in the use of microteaching have been explored by Carver and Wallace (1975), who combine micro and peer teaching; by DeLorenzo (1975), who found that experience with microteaching reduced anxiety in student teachers; and by Dugan (1967), who developed an intensive summer institute around microteaching. Forrester (1974) and Moore (1977) provide tips for using the demonstration lesson in which trainees observe and critique a lesson as it is being taught.

Implicit in most articles on teacher training is the assumption that the trainees are either full-time students or that they are a relatively homogeneous group in terms of teaching experience, prior training, career plans, etc. Unfortunately, such assumptions cannot be made about the students attracted to the TESL methodology course at the University of Colorado at Denver. Part II provides a description and explanation of a one-term course offered at UCD for a diverse population of non-traditional TESL students.

3. The one-term TESL course

Need. It is, of course, every teacher's fantasy to have relatively small, homogeneous classes. Events of the past few years appear to have made such fantasies even less likely than before. With the recent influx of non-English speakers to urban areas, many non-ESL teachers find that they need ESL skills to work effectively with their pupils. As a result, ESL teacher-training programs are increasingly attracting non-degree students seeking a one-term, all-purpose course which they hope will prepare them to work with the non-English speakers in their classes. This group on "non-traditional" TESL students includes bilingual education teachers, paraprofessionals, adult basic education volunteers, public school teachers with non-English speaking pupils and an occasional walk-in, that is, an interested undergraduate or special student. These non-traditional TESL students bring a wide variety of knowledge, skills and experience to the TESL classroom; some are experienced teachers in another field, others have never taught; some have a thorough understanding of linguistics and English structure, others have none; some have worked extensively with people from other cultures, others have no cross-cultural experience; some speak several languages other than English, others are monolingual English-speakers; some are interested working with children in "traditional" classroom settings, others are interested in the "non-traditional" education of adolescents or adults. The list could be extended indefinitely. It is a curious and confusing array of people, but for teacher training programs in large urban areas, this is the reality of higher education.

The dilemma which teacher trainers face in this situation is, of course, classic: How does one give the trainees techniques they can use and the theoretical information required to make the most effective use of those techniques? The methodological framework described below was developed in response to precisely this question. Although developed for use in a college or university setting, we believe it is a format which, with a few adjustments, would prove ideal for the inservice training needs of practicing teachers. The priorities for the course, in order, are:

1) to provide students with a bag of tricks, a number of classroom techniques which they can use immediately.
2) to provide students with the ability to adapt material from a variety of sources and to write their own materials.
3) to provide students with the opportunity to observe and analyze master teachers of ESL in a variety of classroom situations.
4) to provide students with the opportunity to practice a variety of techniques.
5) to provide students with an understanding of the resources available to

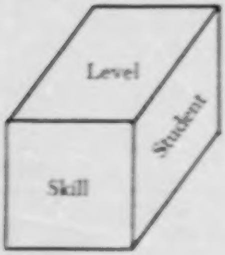
them (i.e., professional organizations, professional literature, local resource centers and libraries, materials available, etc.).

6) to provide students with sufficient theory to understand the implications of the choices they make.

The Confounding Cube. Given these general priorities, the trainer must develop a syllabus which systematically addresses the needs of all the trainees. These needs, in terms of the context in which trainees might be teaching, can be represented graphically by the Confounding Cube of Fig. 1.

Figure 1

The Confounding Cube



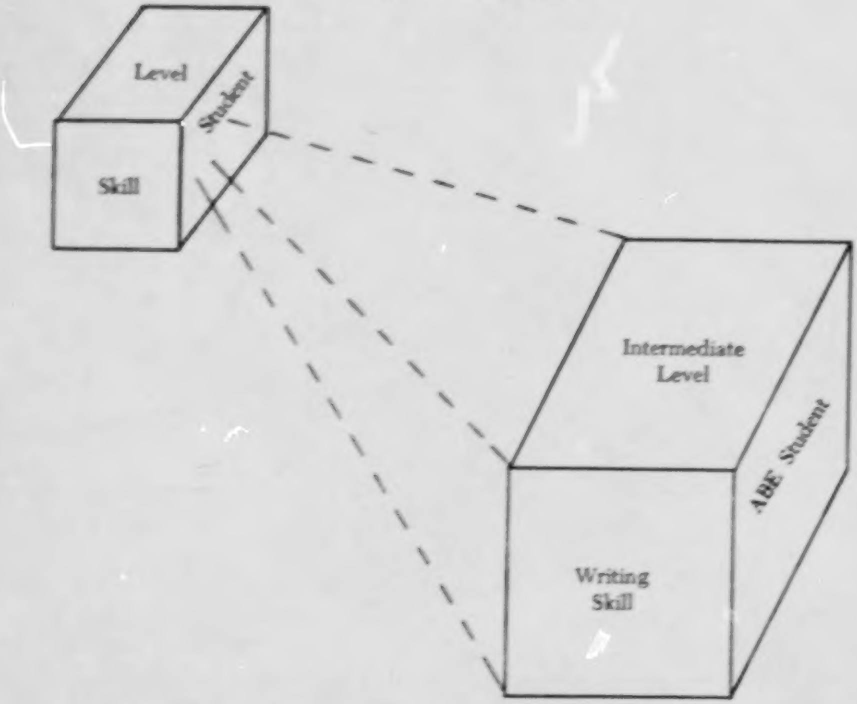
1. Level
    a. Beginning
    b. Intermediate
    c. Advanced
2. Skills
    a. Listening
    b. Speaking
    c. Reading
    d. Writing
3. Students
    a. Public School
        1) Elementary
        2) Secondary
    b. Intensive Course
    c. Adult Basic Education

In teaching one-term methods courses to heterogeneous groups of trainees, teacher trainers must be aware of the diverse challenges which the trainees will face in the field. The various combinations of levels, skills and students—represented graphically by the Confounding Cube—must be systematically addressed in the methods course.

Because the trainer cannot assume that everyone in the class will be teaching in the public schools, or in an intensive English center, or that they will be focusing on oral/aural skills or that they will be working with students of the same ESL proficiency, s/he must attempt to systematically touch on as many combinations of the cube as possible. Used conscientiously as a framework for preparing lessons, the cube enables the trainer to avoid the oft-cited (for example, Orwen 1971, Paulston 1973) weaknesses of Methods courses: that they represent the experience and expertise of the trainer more than a systematic review of possible methods and techniques. Using the cube in this fashion, the trainer makes explicit the applicability of a particular technique. Figure 2 shows one possibility.

Trainees for whom such a lesson is appropriate would benefit immediately from the presentation while others would be aware of the adjustments they would need to make to effectively use the technique in their classrooms. Such a presentation is especially valuable for the neophyte who begins to acquire an appreciation for the complexity of the field and an understanding of which techniques are appropriate for which student population. The cube also permits the

Figure 2
Focus of a Sample Lesson



An example of how the Confounding Cube is used to focus a presentation; this lesson would present a technique for teaching writing to intermediate-level ABE students. Trainees interested in, or working with, different students or at different levels would be expected to make the necessary adaptions to the technique.

trainer to check him/herself in planning the course to meet the needs of all trainees in the class.

Three component system: Many authors have called for less theory and more technique in TESL methods courses (see, for example, Bowen 1967, Alatis 1974, Lee 1974, Strevens 1974, Schmieder, et al. 1977, Taylor 1978). Arguments for such an emphasis are compelling; i.e. that in short-term training, it is important to give trainees something they can use immediately; that it is the responsibility of a trainer to enter a course with an open mind, prepared to give trainees what they need rather than what s/he happens to enjoy teaching, etc. As Lee (1974:37) states, "getting down to the grass roots in the training of . . . teachers means getting down to the detail of what they have to teach, but not necessarily to the detail of any underlying theory." Bowen (1967:32) supports this perspective: "Short term training should . . . limit itself to surface structure, to teaching the application of the best and most productive of current teaching techniques presented as effectively as possible."

Of course, none of these authors advocate the elimination of theory from methods courses, but their experience has led them to the conclusion that a course which emphasizes the practical over the theoretical will be the most effective. Unfortunately, most writers offer very little specific information on how a trainer is to integrate theory and practice. Most would undoubtedly agree with Lee (1974:37) "General principles, if it is possible to pursue and find them, are better glimpsed and then, perhaps, tracked down through the detail (of teaching techniques) met with rather than received as announcements." However, the three-component framework presented here is, to our knowledge, the first detailed attempt to show how theory and technique might be effectively integrated in a short-term methods course.

The three components are Technique, Practice, and Theory.[3] They are presented to the trainees at the beginning of the session as the framework within which everything in the course will take place. It is made clear that: 1) the highest priority for the course is for trainees to acquire a bag of tricks, a portfolio of techniques for use in their classes; 2) most of the techniques presented will have to be modified somewhat for their use; and 3) an understanding of the theoretical basis and implications of a technique is essential if they, are ESL teachers, are to be efficient, effective decision makers rather than mere performers" (a la Jarvis 1972).

With the above information as background, the trainees are then presented with a series of techniques which they practice and then examine from a theoretical point of view. From the trainer's point of view, there are a number of options available under each component. For our purposes, a technique is a specific classroom activity for teaching or testing a language skill, or grammar/vocabulary point. The essence of the Technique component is its brevity and the explicitness with which the behaviors of the teacher and the students are explained and demonstrated. Practice in this context is any activity which provides trainees contact with the real world; it might take the form of observation of a master teacher, microteaching, group discussion, individual adaptation, etc. The Theory component provides an opportunity to step back from the technique, to establish the context in which the technique might be used, to explore the implications of the technique for the students, the syllabus, and the teacher. This might take the form of class or small group discussion, a mini-research paper, or a reading assignment. Chart 1 provides a picture of the options available to the trainer.

In our experience, the most effective use of time provides for approximately 50% for technique, 25% for practice, and 25% for theory. These percentages would apply both to individual lessons and to the course as a whole. Thus, the course is developed around x number of technique presentations with time structured so that trainees have the opportunity to work with the techniques and to reflect

---

[3] Theory is used here to mean a rationale which guides classroom practice; theory encompasses approach and method as defined by Anthony (1963).

CHART 1
A 3-Component Model for Teacher Training

Technique	*Practice*	*Theory*
Presentation by trainer	Observation of class	Class or small-group discussion
Demonstration lesson	Small-group discussion	Outside reading
Micro-lesson	Individual adaptation	Mini-research project
Shock language lesson	Lesson-plan development, adaptation	Trainer-facilitated discussion
Film, filmstrip		
	In-turn demonstration	
Slide presentation		
Video tape	Case-study problem solving	

upon the theoretical implications of the techniques. (It is important to keep in mind that each technique is presented with specific reference to a particular skill, language competence level, and student population. Thus, it is necessary for the majority of trainees to adapt the technique before it will be of use to them.)

At this point, the astute reader will raise objections: There is nothing new here. What makes this different from the average methods course? How does this course avoid becoming merely a series of ESL workshops or inservices? In terms of content, there is nothing new here, and an observer might not be able to distinguish this course from any number of TESL methodology courses taught around the country. And the major pitfall of this approach is, of course, that the series of techniques will become merely that, a series of techniques, rather than pieces of a mosaic which acquire meaning as part of a whole.

The uniqueness of this program can be evaluated from the perspective of the trainer and the trainees. The trainer is given a framework which permits him/her to systematically relate theory to practice. Although the focus of the course is a series of techniques, this is only the surface structure of the class. In fact, in planning lessons, the trainer first establishes the theoretical issues to be addressed and then selects the technique(s) which will illuminate those issues. The effect is that the techniques are foregrounded, but there is always the partially hidden agenda of the theory which they exemplify. The trainer thereby avoids the temptation to dwell on theory to the detriment of practice or to overemphasize pet techniques.

From the trainees' point of view, the course provides a selection of specific classroom techniques which are immediately available for use. In addition, the trainees acquire the essential skill (which is not always specifically taught) of adapting techniques and materials for their own use. And, finally, the trainees acquire, bit by bit, a theoretical perspective on what they are doing in the classroom. The constant insistence of the trainer to establish a link between techniques and theory results (hopefully) in the trainees' appreciation of the inseparability of the two.

4. Course outline and lesson plan

Appendices I and II contain a sample course outline and a lesson plan developed according to the above framework. The syllabus has been adapted slightly from the course outline presently being used in a methods class at the University of Colorado at Denver; it is designed to cover a fifteen-week semester of classes which meet once a week for a three-hour session. The syllabus is presented as one example of how the above model might be implemented; obviously, the needs and experiences of other trainers will dictate modifications before the syllabus could be used elsewhere.

It is important to stress that preparation of the course (and individual lessons) proceeds from theory to technique. That is, the trainer first develops a list of theoretical issues that s/he wishes to cover in the course of the semester and then selects specific techniques which exemplify those issues. However, a major part of class time and virtually all of the assignments focus on classroom technique.

It will be noticed that Rivers and Temperley (1978) is shown as the text for the course. There are, of course, a number of good texts which might be used for example, Finocchiaro (1974) and Paulston and Bruder (1976). In addition to the text, we have suggested a number of articles which focus on specific teaching and testing techniques. We have kept the reading to a minimum because a major assumption of the course is that trainees will be spending a considerable amount of time searching out techniques which are relevant to their needs. We felt, therefore, that an extensive required reading list would preempt their flexibility in finding resources of specific value to them.

One further comment about the course outline: it should be noted that special attention has been given to the three areas of skills, level, student characteristics of the "confounding cube" in developing the sequence of issues and the presentation of techniques. The sequence in which ESL skills are addressed is probably fairly standard for most methods courses. With regard to techniques, we have attempted to provide a balanced presentation of techniques which would be appropriate for beginning, intermediate, and advanced ESL students in programs at public schools, intensive English centers, and ABE programs. The proportion of time that a trainer devotes to any of these areas would depend, of course, on the particular needs of the trainees in the class.

The lesson plan is presented as a specific example of how the framework developed in this paper might translate into the classroom. The amount of work is ambitious but the uniqueness of the approach is in the presentation of as many techniques as possible in the time allowed. If it becomes necessary to reduce the amount of material covered in class, we generally choose to sacrifice the theory section. Discussion of particular techniques often overlaps into subsequent sessions and, if a particularly important issue requires more time, a theoretical question can be assigned for trainees to ponder for the next session.

In our experience, the emphasis on technique rather than theory has made it possible for all students to benefit from class sessions. Trainees with ESL or

L2 teaching experience generally contribute unexpected insights to class discussions, and they often ask questions which force the discussion to a higher theoretical plane. Neophytes to ESL, on the other hand, frequently ask extremely naive questions, whose value lies in forcing the trainer and more experienced trainees to carefully evaluate very fundamental assumptions about L2 teaching. This fruitful interchange of ideas flows directly, we believe, from the minute examination of techniques; if the focus were allowed to shift from technique to theory prematurely, the less experienced students would not be able to contribute to or benefit from the discussion.

It should be noted that a significant portion of each class session is devoted to adaptation and demonstration of techniques by the students. It might be that one of the techniques is initially demonstrated by a trainee, or that a trainee is required to demonstrate a technique introduced by the trainer. Also, we have had success with small-group demonstrations; trainees seem less hesitant to adopt student/teacher roles in front of fewer people.

Regardless of how a technique is presented and practiced, trainees are consistently required to generate technemes (Stevick 1972a) for improving the rhythm of class activity. Technemes are small variations in a technique which result in improving the appropriateness of the technique for the learners (either by challenging the faster pupils or by helping the slower pupil). Students are also required to evaluate and adapt techniques and materials according to Stevick's (1972b) criteria of strength, depth and lightness. The intended effect is that, although only five or six techniques may be presented in one lesson, each student will have produced meaningful variations of each technique for his/her own use.

In conclusion, we cannot claim that the framework presented here provides the ultimate solution for teacher trainers who are faced with a bewildering array of trainee expectations for the one-term methods course. The demands of EFL/ESL instruction are far too complex for "ultimate solutions." However, we can report that the approach outlined in this paper has increased the general level of satisfaction with the one-term course. The focus on practical techniques, balanced by the trainers' consistent efforts to relate these techniques to theoretical issues, has produced an experience in which trainees acquire skills which they can use on Monday morning coupled with an appreciation for the theoretical foundation on which the techniques rest.

APPENDIX 1

SAMPLE COURSE OUTLINE

Course Requirements:

1. Six mini-reports on specific classroom techniques. Reports are to be written so that another individual could read the report and then use the technique without further explanation or clarification. Reports are to be typed, organized according to a standard format and submitted in duplicate.
2. Classroom demonstration of one technique. Demonstrations are to be brief but faithful replications of one of the techniques presented in the mini-reports. Unless

the trainee makes other arrangements, the rest of the class will serve as ESL students for the demonstration.

3. Six hours of observation of master ESL teachers. Reports on the observations are to be brief (2-3 pages) and are to focus on the specific techniques used by the teacher. A summary statement should indicate how the techniques could be adapted for other teaching situations.
4. Final examination on readings, discussions. The exam will cover both knowledge and skills acquired in the course.

Course Syllabus:

Session	Issue	Techniques	Reading
1 2 3	Structural Pattern Drills	—Shock lesson (Arabic) demonstrating repetition, substitution, transformation drills. —Dialogue adaptations —Mechanical, Meaningful, Communicative drills.	McCready 1972 Paulston 1972 River 1978, Chpts 1, 4 Stevick 1972, a, b
4 5	Teaching Listening Skills (2 mini-reports due)	—Dictation —Lecture note taking —Total Physical Response —Sound discrimination exercises (from Morley, 1972) —Use of puppets	Morley 1972 Rivers 1978, Chpt 3 Sutherland 1967
6 7	Teaching Speaking Skills (Observation reports-3 hrs. due)	—Minimal pairs —Use of props (mirrors, charts, etc.) —Stress and intonation in drills	Rivers 1978, Chpt 5 Robinett 1972 Strain 1972
8 9	Teaching Reading Skills (2 mini-reports due)	—Language experience —Teaching vocabulary —Using unedited texts —Teaching skimming & scanning	Clarke and Silberstein 1977 Rivers 1978, Chpts 6, 7
10 11	Teaching Writing Skills	—Handwriting —Controlled compositions —"Free" composition techniques —Evaluation techniques	Brière 1966 Kaplan 1968 Knapp 1972 Lawrence 1972, 1977 Rivers 1978, Chpts 8, 9 Raimes 1976 Seward 1972, 1979 Seager 1972
12 13 14	Testing ESL	—Testing oral/aural skills (dictation, interview, oral cloze tests) —Cloze test construction —Constructing distractors for multiple choice tests —Correlating test results with standardized tests	Harris 1969
15	Final Exam	—Requires students to evaluate and adapt specific techniques, materials, tests.	

Course Readings

Allen, H. and Campbell, R. 1972. Teaching English as a Second Language (2nd Edition). New York: McGraw-Hill.

Brière, E. 1966. "Quantity before Quality in Second Language Compositions." *Language Learning* 16.3&4, 141–151.

Clarke, M. A. and Silberstein, S. 1977. Toward a Realization of Psycholinguistic Principles in the ESL Reading Class. *Language Learning* 27.1, 135–154.

Croft, K. 1972. *Readings on English as a Second Language*. Cambridge, Mass.: Winthrop Pub.

Harris, D. P. 1969. *Testing English as a Second Language*. New York: McGraw-Hill.

Kaplan, C. 1968. *Guided Composition*. New York: Holt, Rinehart and Winston.

Knapp, D. 1972. "A Focused, Efficient Method to Relate Composition Correction to Teaching Aims." In Allen and Campbell, 213–221.

Lawrence, M. 1972. *Writing as a Thinking Process*. Ann Arbor: University of Michigan Press.

Lawrence, M. 1977. *Reading, Thinking and Writing*. Ann Arbor: University of Michigan Press.

McCready, G. 1972. "Developing a Lesson Around a Dialogue." In Croft 1972, 106–114.

Morley, J. 1972. *Improving Aural Comprehension*. Ann Arbor: University of Michigan Press.

Paulston, C. B. 1972. The Sequencing of Structural Pattern Drills. In Croft, K. 1972, 148–162.

Raimes, A. 1976. "Composition: Controlled by the Teacher, Free for the Student." in Fanselow and Crymes, *On TESOL 76*, Washington, D.C.: TESOL, 183–194.

Rivers, W. and Temperley, M. S. 1978. *A Practical Guide to the Teaching of English as a Second or Foreign Language*. New York: Oxford University Press.

Robinett, B. W. 1972. "Simple Classroom Techniques for Teaching Pronunciation." In Croft 1972, 80–86.

Seward, B. H. 1972. "Teaching Cursive Writing to EFL Students." *English Language Teaching* 26.2, 169–178.

Seward, B. H. 1969. *Handwriting in English: A Practice Book*. Denver: The Center for Research and Education.

Seager, W. 1972. Classroom Techniques for Controlling Composition. In Croft 1972, 232–244.

Stevick, E. W. 1972a. "Technemes and the Rhythm of Class Activity." In Allen and Campbell 1972, 88–94.

Stevick, E. W. 1972b. "Evaluating and Adapting Language Materials." In Allen and Campbell 1972, 101–120.

Strain, J. E. 1972. "Teaching a Pronunciation Problem." In Croft 1972, 71–79.

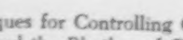
Sutherland, K. K. 1967. "Dictation in the Language Classroom." *English Teaching Forum* 5.2, 10–14.

APPENDIX II

SAMPLE LESSON PLAN

Technique #1: Handwriting Exercises

1. Shock lesson: Arabic script. Trainees are presented with the Arabic alphabet, allowed to struggle for a short time with the symbols (Arabic is written from right to left, vowels are not normally represented in cursive, and the form of many letters changes dramatically when attached to other letters). Trainees are then introduced to the concept of organizing the letters according to distinctive features, and teaching the letters in groups.
2. English presentation: Trainees practice with English graphemes organized on the same model (from Seward 1972, 1979).
   Example: The *e* group.
   . e, i, u, t



3. Discussion, clarification.

Technique #2: Controlled Composition: Sentence Combining.

1. Trainees are presented with exercises in sentence combining, told to work the exercises.

Example: 1. The salesman ate ten sandwiches. The salesman felt sick.
2. I took my umbrella to work. It was raining yesterday.
(From Lawrence 1977, 77)

2. Trainees produce similar exercises from various texts intended for a variety (in terms of both level and age) of ESL learners.
3. Two trainees write excerpts from their adapted exercises on the board and explain (using Stevick 1972b) the importance of their adaptations.
4. Discussion, clarification.

Technique #3: Controlled Composition: Scrambled Paragraphs.

1. Trainees work with scrambled paragraph exercise, attempt to discover optimum sentence order.

Example: The seven sentences below constitute a paragraph, but they have been printed in a disordered sequence. Reconstruct the paragraph. In the space at the left put the number of the sentences in the order in which they should occur.

______ 1. Furthermore, tacos, enchiladas, chili, and frijoles are common fare for western tables.
______ 2. Although the Spanish political and military influence has long since disappeared, some influence remains in the continuing use of many words of Spanish origin.
______ 3. The western and southwestern parts of America were once under Spanish control.
______ 4. Most western Americans know what is meant by loco, amor, and pronto.
______ 5. This infiltration of Spanish into English has added color and a certain special flavor to our speech.

(From Kaplan 1968, 131)

2. Discussion, clarification.

Technique #4: Free Composition: Quantity before Quality.

1. Trainees are introduced to the concept of free association writing, told to write as much as possible in 3 minutes on any topic.
2. Clarification: The rationale (see Brière 1966) is explained, and the technique is explained in detail.

Technique #5: Controlling Free Writing

1. Trainees are presented with a topic, picture, film or filmstrip and given a specific writing assignment (i.e.: Compare and contrast these two pictures; Describe what happened in the film or filmstrip.)
2. After the assignment is completed (5-10 minutes) the trainess read their compositions to the class.
3. The best composition is written on the board.
4. The composition is analyzed for rhetorical devices. Trainees are asked to identify the rhetorical devices which are required by the task (see Raimes 1976).
5. Raimes (1976) is assigned for homework. Trainees are asked to generate assignments which, because of the topic and the assignment, control the ESL students' writing.

Technique #6: Evaluating Writing: Knapp's checklist.

1. Trainees evaluate an actual ESL composition using Knapp's checklist (Knapp 1972; trainees have been required to read Knapp's article before class—see SAMPLE COURSE OUTLINE, Session 10).
2. Group discussion: compare results; discuss rationale for selecting, marking errors; discuss future assignments for the writer of the composition.

Synthesis and Evaluation

The discussion after each technique has been held to a minimum, the primary objective being to clarify the purpose and implementation of the technique. This final synthesis and evaluation is intended as an opportunity to explore the theoretical implications of the techniques and to discuss ways in which the techniques might be integrated into the curriculum.

Trainees are encouraged to relate the techniques to the assigned readings and to evaluate the appropriateness of each technique to their particular situation. The trainer acts as a facilitator, guiding the discussion (as much as possible) to the following points:

1. The place of handwriting instruction in the ESL curriculum. The differences between handwriting needs of children and adults, between literate and illiterate learners, printing first vs. cursive first or vice versa, handwriting problems of specific language groups (Arabic, for example), etc.
2. The theoretical foundations of controlled versus free composition instruction. Strengths and weaknesses of both techniques. Integrating both techniques. The student journal.
3. The need for helping ESL students achieve independence at the paragraph and composition level of writing as well as at the sentence level. The need for ESL teachers to understand the requirements of English discourse, and to consistently shift back and forth between controlled and free composition assignments.
4. Techniques for integrating reading and writing instruction. The pros and cons of using models in teaching ESL writing.

Part V

Research

The Social and Linguistic Behavior of Good Language Learners

R. W. Cathcart, M. A. Strong & L. W. Fillmore

University of California, Berkeley

This paper reports on the individual differences in English language acquisition of 30 immigrant children after one year of exposure to the second language. We focus on the process of language learning as evidenced by the children's behavior in their classrooms and while at play. The relationships of several aspects of cognitive style and personality to the social and linguistic behaviors of the children are also discussed.

An important aspect of the study is that the subjects' communicative competence is assessed not only by standard elicitation techniques (repetition, story-telling and interviews), but also from numerous observations of their participation in regular school activities. These observations allow us to show the relative usefulness of different language sampling techniques, emphasizing in particular the difference between elicited language and that collected from the learner's environment. In addition, we suggest possible implications of the research for teachers of second languages and those who have second language learners in their classes.

Criminals, cultists, and slow language learners have one thing in common; their aberrations are usually explained in the same two ways. Either the system is to blame, or the individual is deficient. Luckily for teachers, they are more often faced with the slow language learners, but unfortunately they have yet to be shown how to determine whether the system or the individual is at fault.

There is, however, a third alternative which is based on the belief that different individuals have different learning styles. Thus slow language learners are not deficient in their abilities to learn, but they exhibit characteristics which inhibit their learning in the particular situation in which they find themselves. The problem for teachers is first to identify the individuals' characteristics and then to guide them towards the most suitable learning program. This paper will summarize some recent research, in particular our own, on learner characteristics and the assessment of their language, and will focus on the implications of this research for teachers. We make the distinction between social styles, which encompass certain personality characteristics, and cognitive styles, which determine approaches to problem-solving.

The forerunners to this research were Carroll (1963), who looked at language aptitude, and Gardner and Lambert (1972), who examined attitudes and motivation. Research on the relationship of cognitive styles to language learning is still relatively recent. Naiman, Frölich and Stern (1975) included certain cognitive styles among a battery of independent variables in their study of the good language learner. They found that field-independence was related to two of their language measures, imitation and listening comprehension, and that tolerance of ambiguity was related to listening comprehension. Their subjects were junior-high and high school students. Fillmore (1976) conducted a longitudinal study of five Spanish-speaking children learning English in an American elementary school. Her observations indicated that several traits appeared to promote language learning. These were: sociability; preference for play situations which required language; flexibility; and a straight-forward approach to problem-solving.

Research on personality factors has yielded contradictory results, partly because the age of the subjects and the language learning environment have varied widely from study to study. The factors which researchers have studied include sociability, extraversion, introversion, empathy, anxiety, ego-permeability and self-esteem. In our study (reported last year at the Second Language Research Forum in Los Angeles), we examined four cognitive styles: field-independence, tolerance of ambiguity and breadth of categorization, and four social styles: outgoingness, self-esteem, submissiveness and egocentrism.

The subjects were twenty children, twelve kindergartners and eight first/second graders, of mixed language backgrounds. They were given standardized and adapted cognitive style tests, they were observed in different settings, and the teachers rated their personality characteristics. Language acquisition was measured from a series of elicitation tasks and observation notes. The elicitation procedures included imitation and storytelling tasks, and a 20-minute interview. The following measures were used in the analysis: errors in imitation, breadth of vocabulary in storytelling and free conversation, and structural complexity from the interview. Additional examination was made of the form and function of the spontaneous classroom utterances.

Statistical analysis of the rank orders on the cognitive and social style measures and language scores produced the results seen in Table 1. The significant associations for the first/second graders were between outgoingness I (as measured from teacher ratings) and structural knowledge; outgoingness II (as measured from the observational data) and all language measures; and adaptive flexibility (measured by path-finding task) and all three language measures. For kindergartners, there were significant associations between adaptive flexibility and structural knowledge, and between field-independence and vocabulary breadth.

Separate statistical analyses were run on the two groups of subjects for two reasons. First, the class structures at the kindergarten level were so different from those of other grades that the extent of permissable student interaction varied considerably, thereby confounding observations on outgoingness across grades. Secondly, raw scores on the adapted tests are not comparable across such wide age ranges.

The results do seem to indicate differences across the two levels. Although no precise conclusion can be drawn from results based on so small a sample, tentative but worthwhile inferences can be made, which may form the hy-

TABLE 1

Kendall's Tau Values of Association Between Three Language Measures and Nine Personal Style Variables.

	Language Measures	Personal Styles								
		Cognitive Styles				Social Styles				
		1	2	3	4	5	6	7	8	9
Gr ½ N = 8	Struct. Knowledge	0.04	0.25	0.71*	0.46	0.71*	0.89*	−0.43	−0.24	0.24
	Imitation	−0.21	0.07	0.75*	0.36	0.52	0.71*	−0.62	−0.24	0.24
	Vocab. Breadth	−0.14	0.21	0.75*	0.36	0.52	0.71*	−0.62	−0.24	0.43
Kg. N = 12	Struct. Knowledge	0.35	−0.23	0.41*	0.07	0.38	−0.02	0.18	−0.18	0.08
	Imitation	0.28	−0.26	0.02	0.18	−0.03	−0.15	0.06	0.06	−0.17
	Vocab. Breadth	0.43*	−0.24	−0.02	0.22	0.36	0.15	−0.21	0.02	0.30

1 = Field Independence
2 = Tolerance of Ambiguity
3 = Adaptive Flexibility
4 = Breadth of Categorization

5 = Outgoingness I
6 = Outgoingness II
7 = Self-esteem
8 = Submissiveness
9 = Egocentrism

potheses for our future studies. It would appear that good language learners at the first and second grade level are outgoing and flexible. The inconsistent results for kindergartners may be partially explained by the well-known unreliability of trait-measurement in such young children.

On the other hand, it is interesting to look at the results which were significant for kindergartners. One possible interpretation is that different personal styles influence different kinds of acquisition at the kindergarten level. Thus it seems that a kindergartner who is flexible will more quickly acquire structural knowledge, and one who is field-independent will use more novel vocabulary. This is logical since the disembedding skills associated with field-independence may relate to the ability to isolate and identify single words, while flexibility, or the ability to generate novel solutions to a problem, could well be related to the creation of new patterns of words. While this explanation seems plausible, it remains to be established why the same findings were not obtained for the older children.

One possible source of the differences between grade levels might be the activity structures of the classrooms themselves. In the kindergarten classrooms many of the activities were group projects, where noise level was higher and children were permitted to interact more freely. On the other hand, in the first and second grade classrooms much of the work was individualized and children were expected to function more on their own. The heterogeneity of the tasks encouraged less cooperation and thus less interaction. In this type of environment there is more responsibility on the individual to create interactive situations for herself, which makes outgoingness an essential quality if the child is to obtain plenty of language input from her peers.

Analysis of the language samples revealed two important factors for consideration. The first was that individuals frequently performed differently on different skills. That is, no two rank orders of language measures correlated for the kindergartners. Thus a child's score on one measure may not be an accurate representation of her overall skill. This is illustrated by the rank order of two subjects, seen in Table 2. If these two children had been tested only on imitation, they would both have appeared good language learners. If they had been tested only on storytelling, they would have been judged poor language learners. However, their ranks on structural complexity give yet a third picture. Imelda appears proficient while Susanna is ranked the lowest of the 12 kindergarten subjects. Furthermore, ESL teachers rated Imelda one of the most advanced ESL learners at her school, and Susanna was ranged 15th our of 20 at hers. This gives rise to the interesting possibility that children's learning and interpersonal styles might affect language acquisition, not necessarily in a general way, but differentially, so that the learning of some skills may be enhanced and of others retarded by the same behavioral or cognitive qualities.

TABLE 2

Rank Orders for Two Subjects of Three Language Measures

Measure	Imelda	Susana
Imitation (errors)	3	1
Story (vocabulary)	10	12
Interview (structure)	2	12
N = 12		

The second factor for consideration was the qualitative difference between the natural language samples and the elicited language. Vocabulary breadth corresponded in both samples, but the form and function of the language varied so greatly that other comparisons were difficult. A typical example of the difference between the story, the interview, and spontaneous language is given in the Appendix.

It is important here to consider the influence of context on language. Fillmore (1976) has pointed out that in the early stages of second language acquisition, children will get themselves into contexts, or situations, which involve them in language. She formulated these social strategies, among others:

> 1. Join a group and act as if you understand what's going on, even if you don't.
> 2. Give the impression, with a few well-chosen words, that you can speak the language.
>
> (pp. 667-669)

It is not surprising that the language which results from a social context where these strategies are employed is very different from the language used in a regulated context where an experimenter is saying: "tell a story", or "answer these questions".

The spontaneous language data revealed interesting individual differences related to language fnuction. In general, 63% of the spontaneous language consisted of directives of some kind. Table 3 provides a breakdown of these directives for the kindergarten children. It can be seen that simple imperatives (e.g. "give me that!") comprise about 25% of the language, non-imperative forms (such as "I need a red" or "that one's mine") make up another 20%, and atten-

TABLE 3

Functions of Classroom Language for Kindergarten Subjects

Subject	Imperative	Non-imp. direct.	Attention	Dispute	Self-verb.	
Cho	5	7	–	3	5	
David	2	6	1	3	4	
Marita	5	1	–	3	1	
Ana Lucia	4	–	7	1	1	
Alberto	3	1	3	–	–	
Sabina	1	3	1	–	4	Other = 3
Jo Sun	2	–	–	2	1	
Elena	1	–	4	1	–	
Juan	–	1	1	2	1	
Imelda	–	–	1	1	–	
Susana	–	–	–	–	–	
Raymon	–	–	–	–	–	
	23	19	18	16	17	T = 95

tion-calling-devices (or ellipted directives such as "hey!" or "Joe!") comprise a further 19%. An additional 17% is language involved in some sort of dispute (such as "I'm not supposed to . . ." or "teacher said so"). Of the remainder, most language consisted of self-verbalizations (17%).

The individual differences in usage seem to be related to stages of language acquisition proficiency. At the earliest stage the child seems to use attention-calling devices regardless of their appropriateness. One child, in an attempt to engage his neighbors in conversation, resorted to pointing and calling attention to objects around the room. His efforts were unrewarded on that occasion, and he was generally ignored. At the intermediate stage similar language is employed, but this time more selectively and appropriate to the context. We observed this from a child who was trying to take part in an ongoing discussion at her table. She nodded and smiled, and when one child punched another, she punched the first. When the conversation turned to broken crayons, she shouted argumentatively, "Me not broken", while holding up one of her crayons. At the more advanced stage many more indirect directives are used, and the children seem to be able to use language according to the demands of the task. In one example, the subject was doing a puzzle with another child and gave him a string of indirect orders of the type: "that belongs to the bee" and "that don't goes here". Seventy-two percent of such indirect directions from the kindergarten group came from the two best language learners. A possible explanation for these apparent stages of language use might be that they are related to personal styles. Another study is needed in order to establish whether all learners go through each stage, or whether the faster learners begin at stage two or three.

The implications of this research for teachers arise from two separate issues. The first concerns the application of the idea that social and cognitive styles affect rate of language acquisition, and the second concerns the definition and assessment of language competence.

For example, if we accept that outgoingness is a desirable quality for language learners, the teacher has two possible courses of action for promoting language learning for the less outgoing children. S/he might try to encourage peer interaction that the children have difficulty initiating on their own, by assigning them more group tasks. In ESL classes this might involve the enlistment of native speakers as coworkers on language learning projects. Alternatively, the teacher might try to match the instructional method with the children's individual styles, providing the resources are available. Thus, the less outgoing child might prosper in the structured environment of ESL classrooms, where there are no demands on her to interact with her English-speaking peers. The outgoing students may do better in immersion or regular classrooms.

For the children who lack flexibility, the teacher again has, logically, two approaches. The first would be to provide language tasks which train the student to be flexible such as finding multiple meanings for words or expressing the same ideas with different constructions. The second method would involve concentrating on pattern examples for these children while providing more open-ended types of exercise for the more flexible learners. Of these two possibilities, it would seem that adapting the program is more plausible than trying to change a child's cognitive style.

The implications for language assessment involve, first, the danger of measuring only one skill. The example referred to in Table 2 shows how individuals may vary across skills. If these two children had been assessed from performance on a storytelling task, as is done in many California school districts, Imelda might have been placed in a beginning ESL class, although her other skills were much further developed. Assessment from an imitation task or any error analysis would have ascribed greater proficiency to Susanna than she merited from her performance on all other kinds of tasks.

Secondly, to measure competence from any relatively limited impressions of spontaneous language may be misleading. As Fillmore (1976) pointed out, and as we have noted in this study, a common and effective social strategy for language learners is to join a group and pretend to know what is going on, preferably using a few carefully chosen language formulas. Listeners are frequently mislead into believing that if an utterance is correct and appropriate, the speaker must have successfully learned the words, learned the "grammar", learned the intonation, and figured out when to put them all together. Thus a shout of: "Hey, gimme one of those", may give a false impression of linguistic sophistication, where it merely reflects a careful attention to context and appropriateness. As a result of this strategy, the child may know and use a great many of these formulas appropriately, without being able to construct imperatives or prepositional phrases in an interview, or even without being able to repeat similar phrases in an imitation task. We need to consider, then, for what purpose we are measuring language competence before we choose how to measure it.

At this point it should be made clear that our research is still at too early a stage for us to reach conclusions. Implications for teachers should be stated only with the qualification that further hypotheses need to be tested before the results may be considered firm. However, we are anxious that practitioners should be constantly aware of the work that is being done by researchers, and we feel strongly that feedback from the teachers is crucial for planning future studies.

To conclude, we believe that individuals differ in their cognitive and social skills in such a way that no one language learning situation is suitable to everyone. Individuals differ in their ability to get the required exposure to the language, and in the amount of help they need from others in order to learn. If the language learning environment (of which instructional programs are types) is compatible with the needs of the learner, then learning is facilitated. If there is a mismatch, then learning will be impeded. At present, the instructional needs of non English-speaking students are defined by the fact that they do not speak English, rather than anything more subtle. Thus, students are simply placed in whatever kind of special program is available or being tried out at the moment. All programs work for some students but no programs work for all. This means that the attributes of each individual must be assessed and catered to. If we succeed in identifying these attributes, then at least we will be in a position to help solve the problem of slow language learners, while happily leaving the problem of criminals and cultists to others.

APPENDIX

Language Samples for One Subject (Cho) from two Elicitation Measures and One Classroom Observation.

Storytelling task

wun a car trying to go to somewhere some boys is wu was riding a bike and that man is mad u that boy n when that man go faster an that boy is flying away as he fell down her bike broke an he got fix it n then u fix it then her car was broken n he can't fix it n then he try to fix.

Interview

Interviewer: All right can you tell me what's happening?
Cho: It's night
Int: Uh huh. It's night and who's sleeping?
Cho: Witch
Int: Uh huh
Cho: Jus witch shoes
Int: Uh huh the witch's shoes. OK so the witch is sleeping and where. . .
Cho: He put de broom on di . . .
Int: In the corner, huh yeh, so she's got her broom in the corner, and where are her shoes?
Cho: Down
Int: Down, yeh, under the bed and where has she got her hat?
Cho: Under de, up de bed
Int: On the bed, right.

Classroom observation (doing a puzzle)

Directives: no, the face (put the face in)
take it
stop that
see, see

Dispute: I told you
yes, I can
cuz I wanna play

Non-Imp. Directives:
that one belongs to him
this belongs to the bee
that don't goes there
no, no
that don't fit
let's do this one again

Self verb:
I am
bird goes here
dere
and dis one goes here
I'm done

On Speaking Politely in a Second Language

Robin Scarcella
University of Southern California

An important task in acquiring a second language is learning the rules and mechanisms underlying appropriate use. However, relatively few researchers have been able to identify the specific aspects of communicative competence which our students need to know to communicate effectively. Even fewer studies have examined the problems which our students encounter when using sociolinguistic rules. In light of this gap, our study focuses on one area of sociolinguistics rarely, if ever, investigated: that of pragmatic competence. Our objectives are threefold: (1) to investigate the politeness strategies employed by both L1 and L2 speakers when addressing speakers upward in rank, of equal rank and downward in rank; (2) to identify the politeness strategies which adult L2 performers have difficulty acquiring; and (3) to provide practical suggestions for teaching politeness strategies.

The subjects of our study were adult ESL learners (N = 20) who were asked to participate in role-play situations specifically designed to reveal the subjects' use of politeness features. (Native English speakers served as the control group.) The situations were video-taped, transcribed, and analyzed in terms of those politeness features suggested by Brown and Levinson (1974). The results are significant in that they suggest the specific areas of politeness which deserve greater emphasis in the ESL class.

One of the most important tasks in acquiring a language is learning the rules and mechanisms which underlie appropriate use. (See, for example, Hymes 1967; Grimshaw 1973; and Taylor and Wolfson, 1978.) Paulston (1974), in fact, contends that learning the appropriate forms of social usage of a language is just as important as learning the linguistic forms. As Hymes (1967) argues, communicative competence includes not only the linguistic forms of a language, but also a knowledge of when, how, and to whom it is important to use these forms.

While many second language researchers and pedagogues have emphasized the importance of communicative competence in teaching English as a second language, (for example, Oller, 1973b; Rivers, 1972; Wilkins, 1976; and Halliday, McIntosh, and Strevens, 1964), relatively few have been able to identify the specific components of communicative competence which our students need to know to communicate effectively. Even fewer studies have examined the problems which our students encounter when using sociolinguistic rules. Indeed, there has been very little research on the acquisition of the sociolinguistic-interactional rules which might govern communicative competence. (But, see Dunkel, 1978; Schmidt, 1978, for some current investigation.) Perhaps this is because we in second language have only quite recently turned to both socio-

and psycho-linguistics when characterizing the language acquisition process, (Hatch, 1978; Hakuta and Cancino, 1977).

In light of this gap, our study focuses on one area of socio-linguistics rarely, if ever, investigated for the purposes of second language research: that of pragmatic competence. More specifically, we focus on the use, or perhaps misuse, of politeness features in adult first and second language performance. Although much of the literature centers on what native speakers do in their own language, (Lakoff, 1973; Brown, 1975; Brown and Levinson, 1974; Ferguson, 1976; and Shimanoff, 1977), relatively few studies have been reported on the acquisition of politeness features by second language learners.

Generally speaking, politeness use is motivated by a speaker's desire to save face,[1] her/his own face, or the hearer's face. According to Brown and Levinson (1974), there are two types of politeness: positive politeness which is used to satisfy the speakers' needs for approval and belonging, and negative politeness which functions to minimize the imposition of the face-threatening act. These politeness types may vary according to the nature of the act itself and status which we define here in terms of the social distance and relative power of the speaker and hearer.

In this paper, we examine the effect of status on politeness in adult second language performance. In so doing, we describe and analyze the politeness strategies employed by native (L1) and non-native (L2) speakers when addressing speakers upward in rank, of equal rank, and downward in rank, (hereafter referred to as superior, equal familiar, and subordinate, respectively).

Our objectives are twofold: 1) to investigate the politeness strategies used by adult L1 and L2 speakers, and 2) to identify those strategies which adult L2 performers have difficulty acquiring. In order to achieve these goals, we undertook the small study described below.

1. Method

Subjects. 20 adult L2 performers (10 advanced and 10 beginning)[2] participated in our study. The subjects were male, spoke Arabic as a first language, and ranged in age from 20 to 27. Our control group,[3] consisting of 6 native English-speaking males, ranging in age from 18 to 26, enabled us to compare the speech of adult L1 and L2 speakers.

Task. The subjects were asked to participate in three role-play situations. These situations were specifically designed to reveal the subjects' use of politeness features when speaking to superiors, equal familiars, and subordinates. The situation is described below.

---

[1] Brown and Levinson (1974) define face as a set of wants satisfiable only by the actions (including expressions of wants) of others.

[2] Level was determined by scores on the American Language Institute (USC) Placement Exam, Spring, 1978.

[3] The subjects of the control group were university students who formed part of the USC debate team.

You are planning an office party. You invite your boss, the clerk who works under you, and your good friend, a fellow employee. You request that each of your guests come unaccompanied by his wife.

There were many advantages in using role-play. First, it allowed us to obtain complete conversational interactions, containing both conversational openings and closings. Second, it enabled us to control the conversation to a certain extent. This in turn provided us with comparable samples of speech. Third, it facilitated videotaping, important when examining paralinguistic features of discourse. Fourth, we found, as did Andersen (1977), that role-play speech tapped an awareness of language appropriateness. However, a word of caution seems necessary here. Several investigators have suggested that role-play speech is not representative of natural language. (See, for example, Mitchell-Kernan and Kernan, 1977; and Littlewood, 1975, for role-play in adult L2 performance.) Thus, while role-play speech may reflect one's ability to vary language according to the sociolinguistic context, it probably does not indicate what one actually says or does in more relaxed, natural situations.

Data Analysis. Each of the 78 speech samples was transcribed by the two researchers. The samples were then analyzed in terms of those features of positive and negative politeness suggested by Brown and Levinson (1974). (Measures of politeness were decided upon in advance by the two researchers. See appendix for exact definitions.)

2. Discussion of findings

In this section we examine positive and negative politeness in L1 and L2 performance. Comparisons of the use of politeness features by beginning and advanced L2 performers and native English speakers are made whenever possible.

Positive Politeness. In positive politeness, Brown (1975) states, "H (the hearer) is treated as a member of an in-group, a friend, a person whose personality traits are known and liked." (p. 6). Common ways of achieving positive politeness include: 1) expressing an interest in and noticing H; 2) using in-group language (for example, address terms, slang, expressions such as "ya know", ellipsis, inclusive "we", and prosodic and kinesic features; 3) making small talk; and 4) showing and seeking agreement (Brown and Levinson, 1974).

1. *Expressing an interest in and noticing H.* One of the principal means of expressing politeness is by showing interest in and noticing H. This can be done through greetings. In our data, most of the conversations, (75/78 or 95%) began with greetings. Only in the beginning (low level) enactments were greetings absent. Indeed, we suspect that politeness formulas such as greetings may be one of the earliest means of demonstrating politeness in a second language. Our data do, however, indicate that even advanced (high level) L2 performers have difficulty using appropriate greetings.[4]

---

[4] To facilitate identification of the data, we have labelled our examples. The following abbreviations are used: *NS* stands for Native Speaker, *HL* stands for High Level L_2 Speaker, and *LL* is used for Low Level L_2 speaker. *B* is for *boss*, *E* is for *employee*, and *C* is for *clerk*.

(1) E: Hello
F: Hello. Welcome.
(High Level L2 (HL):Situation 2:Group 1)

2. *Using "in-group" language*. In-group language stresses speaker involvement with the hearer. It is here where we find striking differences between L1 and L2 speakers.

One means of emphasizing friendliness and solidarity is through the use of in-group address terms, (for example, first names) and endearments (for example, *pal, buddy, honey*) (Brown and Levison, 1974). Although the L2 speakers used endearments to stress speaker involvement and minimize imposition (See example (2) below), the L1 speakers did not.

(2) E: But friend, I don't want you to bring your wife.
(Low level L2 (LL):2:7)

Slang, another marker of in-group language was frequent, at least in the L1 data, and was used particularly with equal familiars by the native speakers. The non-native speakers used slang randomly, that is, in all three role-play enactments.

TABLE 1: Slang

Percentage[a] of slang[b] directed to boss (superior), friend (equal familiar), and clerk (subordinate)[c]

Subjects	To boss	To friend	To clerk
Native speakers	23% (18)	41% (32)	35% (27)
High Level L2	40% (2)	20% (1)	40% (2)
Low level L2	——	——	——

[a] The percentage represents the amount of slang used to each addressee

[b] Slang expressions, counted separately by the two researchers, were defined as expressions of informal language outside of standard language, (e.g., "guy", "Got it!" and "Can ya deal with that!").

[c] The number of occurrences of slang expressions is in parentheses

Expressions such as *ya know, I mean*, and *you understand*, which make the addressee a more active participant in the conversation, are also considered a part of in-group language (Lakoff, 1973). Such expressions appeared frequently in the L1 data, less frequently in the high level L2 data, and infrequently in the low level L2 data. The L2 speakers did not always use these expressions appropriately.

(3) E: Will you almost uh make my home as your home you know.
(HL:2:2)

Ellipsis (here defined as the omission of a subject and/or verb in an utterance), is another marker of in-group language. Ellipsis was present in the L1 data in all contexts. The L2 speakers used less ellipsis, and the type of ellipsis which they employed differed from that used by the L1 speakers. For example, the ellipsis most frequently used by the L2 speakers consisted of phrases used in response to a question.

(4) B: At what time is the party?
E: About 7:30
(HL:1:3)

The absence of ellipsis in the L2 data often made the speech seem overly formal. Consider, for instance, examples (5) and (6):

(5) E: Excuse me sir I am sorry to interrupt you at this time
(HL:1:4)
(6) C: What about the address
E: The address is 2nd and Vermont
(LL:3:10)

Inclusive "we", the *we* which includes both the speaker and the hearer in the conversation, also occurred frequently in the L1 data, mainly with the superior and equal familiar, but also, to some extent, with the subordinate. In contrast, inclusive *we* was absent in the data of the L2 speakers. This finding is significant in that it indicates that inclusive *we* may be acquired quite late in adult L2 acquisition.

Table 2: Inclusive "We"*
Percentage and number of occurrences of inclusive "we" used with each addressee

Subjects	To boss	To friend	To clerk
Native speakers	38% (8)	43% (9)	19% (4)
High level L2	——	——	——
Low level L2	——	——	——

* Inclusive "we" includes both the speaker and hearer in the conversation. (e.g., "I don't understand what we're going to do", "we'll be out of there by 12:00).

In-group language can also be characterized by its prosodic features, (for example, volume and rate). We found that L1 speakers spoke faster to the friend than the superior or subordinate. (See Table Three below.) Thus our data are consistent with the findings of Brown and Levinson (1974). Not surprisingly, the low level L2 speakers, unlike the high level L2 speakers, were unable to use fast speech to convey friendliness. More observationally, we found that both L1 and L2 speakers spoke louder when speaking to subordinates or superiors. And, we noted that generally the high level L2 speakers spoke much

TABLE 3: Speech Rate
Words per minute* used by L1 and high and low level L2 speakers when addressing speakers of different statuses

Subjects	To boss	To friend	To clerk
Native speakers	162	174	156
High level L2	120	182	144
Low level L2	144	120	120

*Speech rate was measured in words per minute, based on the first occurrence of fluent speech in each sample. (All full words, but not word fragments were counted.) We defined a word as a standard orthographic unit. While proper names and exclamatory expressions counted as one word, concatenate forms such as "kinda" were counted as two words.

louder than either the low level L2 speakers or native speakers, (at times, in fact, almost shouting).

Perhaps the most salient markers of in-group language are kinesic. Both L1 and L2 speakers used body gestures which emphasized solidarity. When speaking to an equal familiar, for example, the speakers often leaned forward, stood close, and faced their conversational partner squarely. In addition, there was frequent smiling and laughing, head nodding of approval, touching (especially hand shaking and shoulder slapping at the onset of the interaction), and outward hand and arm movements. When speaking to a superior, on the other hand, there was more distancing, less laughter, inward hand and arm movements (including slumped shoulders), and downward eye gaze. When comparing the L1 and L2 speakers, we observed that the L2 speakers used more outward hand and arm movements, more touching, (especially hand-shaking), less distance and more direct eye gaze. While these kinesic features may indicate friendliness and solidarity to Arabic speakers, English speakers may perceive them as over-friendly and even imposing.

3. *Making small talk.* One way to claim common ground with an addressee and avoid abruptness is by making small talk. In such a way, the speaker enhances the hearer's positive face. In general, the speech data to friend was marked by the presence of small talk in all speech samples. There were, nevertheless, distinct differences in the use of extended openings and closings by L1 and L2 speakers. (We include openings and closings in this discussion since, like small talk, they function to eliminate abruptness in a conversation.) When speaking to a superior, we observed, the L1 speakers only used short openings, such as attention-getters.

(7) E: Mr. Jones. Mr. Jones.
B: Yeah.
E: Uhm. I was wondering uh we're having a party. . . .
(NS:1:2)

Longer openings were reserved for equal familiars and subordinates. The L2 speakers, particularly the low level ones, however, often used extended openings in all contexts, as in (8).

(8) E: Hi Saled.
B: Hi
E: How are you doing
B: Fine just fine
E: What's new.
Uh I am very glad to see (to tell you) we have a party
(LL:1:9)

Ferguson (1976), in discussing politeness formulas, notes that elaborate greetings are often used in Arabic. It may be that the extended openings used by the L2 performers result from a transfer of Arabic discourse-politeness strategies into English.

Closings also differed among the three groups of subjects. The *L1 speakers* used a variety of pre-closings, (such as *alright, okay, yeah*, and *fine*), to politely indicate that the conversation was ending. The L2 speakers, on the other hand, did not. Only in a few cases, in fact, were the high level students able to use the single pre-closing, *okay*. Thus, our data suggest that okay may be the first pre-closing acquired by some L2 performers.

4. *Being agreeable*. Positive politeness can also be achieved by being agreeable. Linguistic markers of this category include repetitions of all or part of what a preceding speaker has just said and positive back channel cues. (Here we define positive back channel cues as markers of agreement such as yes, true, and really.) Repetition in this study was not context sensitive. That is, all of the L1 speakers employed repetition randomly, in all contexts. Repetition appeared to a much lesser extent in the L2 data of the low level speakers. In addition to using repetition, the L1 speakers used a large variety of positive back channel cues in all contexts. Unlike Lakoff (1973), however, we did not find these markers sensitive to the sociolinguistic context. The L2 speakers also relied heavily on these expressions. However, their back channel cues consisted primarily of the expressions *yeah* and *okay*. And, they did not always know the co-occurence rules involved in using these rules appropriately.

(9) F: You're my friend.
E: Okay and please don't uh bring your children and your wife.
(HL:2:2)

5. *Seeking agreement*. Seeking agreement by using such question tags as *right?, okay?*, and *alright?* is also a part of positive politeness. Such question tags were present in all contexts in the L1 data, but lacking in the L2 data.

Negative Politeness

Having discussed positive politeness, we now turn to an examination of negative politeness. Underlying negative politeness, are strategies geared towards preventing infringement upon the hearer's freedom of action. They are central to deferential behavior when addressing those higher in rank and characteristic of social distancing behavior in general. Negative politeness strategies most commonly found in our data were: 1) hedging; 2) indirectness; 3) impersonalizing; and 4) using deferential address terms.

1. *Hedging*. Hedges have a variety of surface manifestations. Among the most common are particles such as *just, maybe, kind of*, and *sort of*, expressions such as *I wonder if, the thing is, something like that*, and to a lesser extent verbs such as *think* and modals such as *could*.

(10) E: I kind of have to talk it over with everybody
(NS:1:2)

(11) E: Just a bunch of the guys getting together
(NS:1:2)

(12) E: We were wondering if you wanted to come along
(NS:1:1)

(13) E: We are thinking about since most of the guys are gonna be there we could invite you
(NS:1:1)

Overall, hedges were much more characteristic of L1 speech than of L2, as illustrated in Table Four below.

TABLE 4: Hedges

Percentage and number of occurrences of Hedges used by L1 and high and low level L2 speakers when addressing superior, equal familiar subordinate

Subjects	To boss	To friend	To clerk
Native speakers	51% (50)	20% (20)	27% (27)
High level L2	52% (9)	5% (1)	41% (7)
Low level L2	21% (7)	43% (14)	34% (11)

In marked contrast, L2 speakers hedged to a very limited extent and low level students especially, did so indiscriminately. Hedges in L2 consisted mainly of *just* and *maybe*. Moreover, there were cases in which these hedges were used altogether inappropriately.

(14) E: I think we have a party tonight.
(HL:1:5)

(15) E: There's maybe something you don't like because maybe you can't bring your wife
(LL:1:7)

Such uses of hedges are clearly inappropriate for the successful completion of acts. Consequently, in addition to lacking a complete repertoire of hedges, the L2 speakers in our study lacked both a knowledge of their semantic function and distribution.

2. *Indirectness*. If as Searle (1969) claims, "in directives, politeness is the chief motivation for indirectness", we might expect more indirectness occurring with superiors and subordinates than with equal familiars. The data in Table Five clearly indicate that in L1 speech, at least, this is the case.

TABLE 5: Indirectness*

Subjects	To boss	To friend	To clerk
Native speakers	46% (12)	34% (9)	23% (5)
High level L2	33% (3)	33% (3)	33% (3)
Low level L2	36% (4)	36% (4)	27% (3)

* Indirectness is considered "not making all of the speaker's beliefs or desires explicit" (Shimanoff, 1977, p. 239). Indirectness was counted independently by the two researchers.

The examples below illustrate the extent to which indirectness varied according to each situation. Example (16) was used with the superior, example (17) with the subordinate and example (18) with the equal familiar.

(16) E: We were wondering if you wanted to come along
(NS:1:1)
(17) E: Why don't you jus' come along
(NS:3:;)
(18) E: Come 'n have a beer party at our place
(NS:2:1)

What is most apparent in these examples is that L1 speakers were more direct with the equal familiar than with either the superior or subordinate. Also, (17) differs from (16) in that it contains an imbedded imperative (Ervin-Tripp, 1974) making it slightly less polite than (16). Contrary to Ervin-Tripp's (1974) claim that hints are used more often to familiars, our findings indicated more hints being used to both superior and subordinates than equal familiars. Perhaps hints, more than any other directive type, vary not only with respect to the situation, but also with respect to task. That is, hints will tend to be used for those acts which most threaten face as, for example in (19) and (20) when they are used with the superior.

(19) E: It's one of the guys kind of things
(NS:1:2)
(20) E: It's one of those affairs where it's just the guys
(NS:3:1)

To equal familiars, on the other hand, imperatives seemed to be more common, as in (21).

(21) E: Ya know you can't bring Joan right?
(NS:2:2)

In sharp contrast to L1 speakers, L2 speakers confined themselves almost entirely to statements of personal desire (Mitchell-Kernan and Kernan, 1977).

(22) E: I would like to invite you to a party.
(HL:1:5)
(23) E: I would like you to join us.
(LL:2:3)
(24) E: I want you to come in a party.
(HL:3:1)

We suggest that such expressions as *I would like* may be, for at least some L2 acquirers, formulaic devices not reflective of the L2 speaker's grammatical competence. As with children's "need" statements, these "statements of personal desire" appear to be one of the first directive types of emerge in second language acquisition. The fact that they were used indiscriminately by low level students seems to indicate that they were devoid of any social (distributional) function.

When using request forms, L2 speakers generally were much more direct.

Simple declarative statements were often used by both high and low level students.

(25) E: You can't bring your wife and your children to this party.
(HL:1:1)
(26) E: You can come alone.
(ILL:3:4)

Unlike L1 speakers who used these only to equal familiars, high level L2 students used these with superiors. To both equal familiars and subordinates, however, they used simple imperatives.

(27) E: Don't bring your wife and your children.
(HL:1:1)

The fact that high level students used simple imperatives mainly to equal familiars and subordinates, and declarative statements to subordinates, implies that they did attach some social meaning to these forms.

In contrast, the low level students used imperatives invariably. This supports our claim that in early stages of language acquisition, many adult L2 performers have difficulty using syntactic means to convey politeness.

We also observed that directives were often accompanied by other features marking politeness. We refer to these co-occurring features as "pres to directives" since they functioned mainly to signal that what follows is a directive. Examples of these pres include *please*, *I'm sorry*, *Excuse me*, and performative utterances such as *I tell you*.

(28) E: Please don't bring your wife.
(HL:1:4)
(29) E: I'm sorry you can't bring your wife.
(HL:1:4)
(30) E: I would like to tell you something else uh you cannot bring your wife with you
(HL:1:4)

The results of our analysis of pres to directives appear in Table Six.

TABLE 6: Pres to Directives*

Subjects	To boss	To friend	To clerk
Native speakers	52% (12)	30% (7)	17% (4)
High level L2	42% (12)	31% (12)	28% (10)
Low level L2	36% (16)	31% (14)	31% (14)

* Pres were defined as expressions preceding directives such as "please", "I'm sorry", and "I tell you".

Here we note that pres to directives were used more often by L2 speakers than by L1 speakers. In fact, pres, especially performatives, appear almost exclusively, an L2 politeness strategy. It is our contention, therefore, that pres to directives: 1) compensated for the L2 speakers' inability to use other means of indirectness, and 2) are one of the earliest politeness strategies acquired by adult L2 performers.

3. *Impersonalizing.* The use of passives and exclusive "we" for the avoidance of "I" and "you" functions to distance the speaker and hearer.

(31) E: We're having like a little party at my place.
(NS:1:3)

(32) E: No wives allowed.
(NS:2:1)

Most common among these impersonalizing mechanisms was the use of exclusive we. Passives, on the other hand, were very scarce.

TABLE 7: Exclusive "We"*

Percentage and number of occurrences of exclusive "we" used by L1 and high and low level L2 speakers to addressees of varying status

Subjects	To boss	To friend	To clerk
Native speakers	46% (12)	34% (9)	21% (5)
High level L2	33% (3)	33% (3)	33% (3)
Low Level L2	36% (4)	36% (4)	27% (3)

* Exclusive "we" is the "we" which excludes the hearer in the activity. Only clear cases of exclusive "we" were counted.

The data presented in Table Seven show that L1 speakers employed this strategy most often with the superior and least often with the subordinate.

L2 speakers, on the other hand, rarely employed this strategy. As a matter of fact, most striking among the features distinguishing L2 from L1 speech was the overwhelming use of the pronouns I and you.

4. *Using deferential address terms.* Deferential address terms (for example, title + last name as in Mr. Jones and sir) are also characteristic of negative politeness. While *sir* was used by both L1 and L2 speakers, the L2 speakers did not always know the precise co-occurrence rules for its use.

(33) E: Hi sir.
(HL:3:5)

3. Summary

To conclude, our analysis led us to some interesting findings. First, it suggests that while some politeness features appear to emerge quite early in adult second language acquisition, (for example, pres to directives such as *sorry* and *please*), others, (such as slang, ellipsis, and inclusive we) do not. See Table Eight below for a summary of our results.

Second, the acquisition of politeness forms appears to precede the acquisition of the sociolinguistic-interactional rules and mechanisms underlying the use and distribution of these forms. (Refer to Table Eight above.) That is, adult L2 performers seem to use politeness features before they have acquired their co-occurrence rules and appropriate distribution.

Third, we would claim, based on the L1 data, that the use of politeness

TABLE 8: Summary of the results

Acquisition of Form[a] and Distribution[b] of Politeness Features

	Native speakers Form	Native speakers Distribution	High level L2 Form	High level L2 Distribution	Low level L2 Form	Low level L2 Distribution
Greetings	+	−	+	−	−	−
Address terms	+	−	+	+	+	+
Expressions such as "Ya know"	+	−	+	−	−	−
Ellipsis	+	+	−	−	−	−
Inclusive "we"	+	+	−	−	−	−
Fast speech[c]	+	+	+	+	−	−
Closings	+	−	+	−	+	−
Positive back channel cues	+	−	+	−	+	−
Hedges	+	+	−	−	−	−
Exclusive "we"	+	+	+	−	+	−
Slang	+	+	−	−	−	−

[a] A form was considered acquired when it occurs in the data at least 10 times.

[b] Distribution is acquired if the use of a form varies in frequency (by at least 10%) when addressed to two different addresses.

[c] Fast speech was considered acquired if an increase in the normal speech rate was evidenced (by at least 10%).

TABLE 9

Summary of Results from Analysis of Positive and Negative Politeness Features

Table 9-A: Positive Politeness*

Percentage of positive politeness features used by L1 and high and low level L2 speakers to addressees of varying status

Subjects	To boss	To friend	To clerk
Native speakers	28% (39)	37% (51)	33% (45)
High level L2	36% (7)	31% (6)	31% (6)
Low level L2	27% (2)	43% (3)	28% (2)

* Based on combined total of all positive politeness features

Table 9-B: Negative Politeness*

Percentage of negative politeness features used by L1 and high and low L2 speakers to addressees of varying status

Subjects	To boss	To friend	To clerk
Native speakers	50% (74)	25% (36)	25% (36)
High level L2	39% (24)	26% (16)	34% (21)
Low level L2	31% (27)	37% (32)	32% (28)

* Based on total number of negative politeness features

features varies according to the status of the addressee, but that L2 performers are limited in both their range of politeness features and their capacity to vary their use according to the social context. As noted in Table Nine below, in the L1 data, the total number of both positive and negative features varied across situations relative to the status of the addressee. As expected, L1 speakers used more positive politeness toward equal familiars and subordinates than to superiors. Conversely, they used more negative politeness to superiors than to equal familiars or subordinates. High level L2 speakers, like native speakers, used more negative politeness towards superiors; however, unlike native speakers, they used more negative politeness toward subordinates. (Perhaps this reflects cultural differences in attitudes towards those lower in rank. It may be the case that lower ranked individuals in this culture do not prompt social distancing as they might in other cultures.) Finally, low level L2 speakers showed much less variety in the politeness strategies they used and lacked any apparent distribution of politeness features.

The results of this study, while preliminary, are very suggestive. In order to confirm and extend our findings, studies in progress examine more naturalistic data, data coming from larger numbers of L2 performers of diverse language backgrounds. The effect of first language interference is being examined. In order to make stronger claims, inferential statistics are being used. Such investigation should lead to a better understanding of the politeness features which adult L2 speakers need to know to communicate effectively as competent speakers in their second language.

Definitions

Ellipsis. Omission of subject and/or verb not necessary for understanding of sentence. (EXAMPLE: Talk to you later.)

Exclusive we. The we which excludes the hearer in the activity.

Expressions that make the addressee a more active participant. Expressions which often function as fillers such as ya know, I mean, and I see.

Hedge. A linguistic feature marking the absence of certainty. Included as hedges are such words as if, perhaps, maybe, hesitations and disfluencies.

Inclusive we. The we which includes both the speaker and the hearer in the activity.

Indirectness. Not making all of the speaker's beliefs or desires explicit, or the avoidance of I or you. (See Shimanoff, 1977, p. 239.)

Positive back channel cue. Marker of agreement which affirms the hearer's remarks.

Pre-sequence. Conversational device functioning as a preparer for and introducer to a conversational sequence. (EXAMPLE: Are you doing anything tonight? Would you like to go to the movies?)

Rate. Measured in words per minute, based on the first occurrence of fluent speech in each data sample. (All full words, but no fragments were included.)

Slang. Informal language often considered outside of standard Usage. Slang expressions include guy, What's up?, and Can ya deal with that?

Question tags. One word tags having rising intonation which function as a means of seeking agreement. (EXAMPLES: You live on Adams, right?, I'll see you at eight o'clock, okay?)

Word. A standard orthographic unit including proper names, contractions, and exclamatory expressions, but not concatenate forms such as *kinda* which count as two words.

The Perception of Politeness in English and Spanish*

Joel Walters

University of Illinois

This paper investigates the perception of politeness in English by native speakers of American English and second language learners of 17 language backgrounds. It also presents data on the perception of Spanish by a group of Puerto Rican speakers of Spanish. Those groups judged speech act strategies for requesting (e.g., can, could, will, do, etc.) using the paired comparisons method. The findings showed high correlations in the perception of politeness between male and female native speakers and between native and non-native speakers of English. Greater unanimity and more categorical perception resulted from the non-native and female native speaker's judgments. In the Spanish request forms, politeness on the part of male speakers could not be predicted from female perceptions or vice-versa.

Most of the work in linguistics, anthropology, and sociology on the form and function of politeness in communication finds its roots in the writings of Goffman (1967). Goffman presents his views of deference in the context of a theory of action or behavior in general, not restricting himself to linguistic aspects of interaction or even to communication in general. He defines the concept of deference as "the appreciation an individual shows to another through avoidance or presentation rituals" (p. 77). By appreciation, he means the value ascribed to an individual, e.g. in assigning high status or prestige to that person. By avoidance rituals Goffman means those actions which allow social distancing to be maintained, including the avoidance of certain topics of conversation. Among presentation rituals he includes salutations, invitations, compliments and requests.

In perhaps the most elaborate extension of Goffman's thinking to linguistic aspects of politeness, Brown and Levinson (1978) present an intricate model and a lengthy taxonomy of politeness strategies with examples of linguistic (including pragmatic, lexical, syntactic and semantic) manifestations of those strategies from three very different languages: English, Tzeltal and Tamil. Their politeness strategies include:

1. Notice the listener's interests, wants and needs
2. Use in-group identity markers
3. Seek agreement/avoid disagreement
4. Offer, promise

---

* The author wishes to thank Jennifer Bahowick for assistance and comments in preparing the paper.

5. Give reasons
6. Apologize

The linguistic examples they provide draw most heavily on speech acts and lexical information. For example, in noticing the listener's wants, interests and needs, a strategy discussed is paying a compliment to the listener. Moreover, some of the politeness strategies actually represent names of speech acts (e.g. offers, promises and apologies). The authors conclude finally that "indirect speech acts have as their prime raison d'être the politeness functions they perform" (p. 147).

R. Lakoff's (1973) rules of politeness seem to be based on some of the same assumptions as Brown and Levinsons' work. One of those assumptions is that the relationship between the speaker and hearer governs the linguistic expression of politeness. Lakoff's rules for being polite dictate to the speaker how to act toward the hearer. They tell him/her:

Don't impose
Give options
Make the listener feel good

Lakoff mainly draws upon lexical strategies as examples for these rules. She cites words like *elimination* as the proper way to avoid imposing on the listener and the use of expressions like *y'know* and *I mean* as ways of making him/her feel good.

Fraser (1975) defines politeness as a function which is based on the hearer's perception of an utterance. It can be categorized then as a perlocutionary effect according to Austin's (1962) tripartite division of speech acts. Fraser goes on to discuss the use of various speech act forms, *please* and certain topics as significant components in the hearer's perceptions.

For this research, the importance of syntactic, lexical, prosodic and discourse information as well as a wide variety of non-verbal data (e.g. physical distance between speaker and listener, setting, eye gaze behavior, body movements) in the perception of politeness was recognized. However, the intent of the study was to investigate how much politeness could be squeezed out of speech act strategies alone. In other words, the principal research question was: How do adults perceive the politeness contained in the speech act strategies of request forms? In that light, the following hypotheses were proposed:

1. Males and females will differ in their perception of the politeness of the speech act strategies of English requests.
2. Second language speakers of English will differ from native speakers in their perception of the politeness of the speech act strategies of English requests.
3. Male and female Puerto Ricans will differ in their perception of the politeness of the speech act strategies of Spanish requests.

1. Method

Subjects. Three groups of subjects participated in comparable experiments.

The first group consisted of sixty college students who were native speakers of American English. Half were male and half were female. The second group of subjects consisted of foreign university students of varying (17) first languages. Approximately half were either Chinese or Korean, with the remaining half equally distributed among fifteen other mother tongues. Forty-five subjects were male and thirty were female. All had achieved a high enough score on the TOEFL to gain admittance to the university. They were enrolled, however, in a course in rhetoric, the most advanced level of training in English for nonnative speakers. The third group was composed of ten Puerto Rican speakers of Spanish; five were males and five females.

Materials. In the case of the native English speakers and foreign students, the stimuli consisted of 14 generic request strategies which were equivalent to the semantic form. These were attached to a standardized lexical content to assist the subjects in making their judgments. (Table 1 provides a list of these strategies.) The origin of this particular set of request strategies is a corpus of 256 requests produced by 32 Puerto Rican children in a role-playing experiment with puppets. The 14 request strategies in Table 1 accounted for more than 97 percent of the corpus. Other data gathered through similar experiments with native English speakers and Armenian speakers of English demonstrated that these same strategies accounted for 80 and 90 percent of the corpora, respectively. Currently, ethnographic and ethological work is being conducted to determine whether these strategies are the same ones which show the greatest frequency of use in naturalistic settings.

TABLE 1

Generic Request Strategies (English) and Sentences Submitted for Judgment.

	Strategy	*Sentence*
6	If you can	If you can give me some rice?
13	May	May I have some rice?
7	Could	Could you give me some rice?
9	Would	Would you give me some rice?
11	Do	Do you have any rice?
1	Will	Will you give me some rice?
10	Can	Can you give me some rice?
4	I came to	I came to ask where the rice is.
2	You can	You can give me some rice?
8	You have	You have some rice?
14	Where	Where is the rice?
5	I want	I want some rice?
3	Imperative	Give me some rice.
12	You have to	You have to give me some rice.

The Puerto Rican subjects all judged 14 generic request strategies in Spanish which, like the English ones, were attached to a standardized lexical content (see Table 2). These strategies were extracted from a similar experiment to the one described above which was conducted in Spanish. The 14 Spanish strategies accounted for 97 percent of a corpus of 256 requests.

TABLE 2

Generic Request Strategies (Spanish) and Sentences Submitted for Judgment.

	Strategy	Sentence
1	Usted no podría	Usted no podría darme el arroz?
6	Me puede	Me puede dar el arroz?
4	Puede . . . me	Puede darme el arroz?
14	Usted tiene	Usted tiene arroz?
11	Si me puede	Si me puede dar el arroz?
9	Quiere	Quiere darme el arroz?
2	Puedo	Puedo llevar el arroz?
12	Usted no tendría	Usted no tendría el arroz?
5	Donde	Donde está el arroz?
10	Usted no puede . . . me	Usted no puede darme el arroz?
3	Me da	Me da el arroz?
8	Imperativo	Deme el arroz.
7	Tiene que	Tiene que darme el arroz.
13	Que	Que me de el arroz.

Procedures. After reading and signing a consent form, each subject was provided with a set of instructions. The English instructions read as follows:

> The concept known as politeness as it applies to language use has been defined as a speaker's symbolic subordination to the addressee. That is, the speaker intends to come across as having fewer rights and/or more obligations (i.e. subordinate) than the hearer.
>
> Each row on the next page contains two sentences which express a certain amount of politeness. Your task will be to decide which member of each pair is more polite. You should ignore context as much as possible. Rely only on the words of the sentence to make your decision about its politeness. Some sentences will seem incomplete. You are to treat these in the same way as the others, relying only on what is written to make your decision about the politeness expressed.
>
> Example:
>
> Shut up . . . . ~~Please~~ be quiet

The Puerto Rican group was administered a translated version of these same instructions. In addition, since the experiment was conducted in a group setting with the foreign students, the instructions were read aloud to them by the experimenter. The subjects from the other two groups participated in the experiment individually.

Following Edwards (1957) the method of paired comparisons was implemented to arrive at a scale of deference for the generic request strategies. This method required that each strategy be compared with every other strategy for the purposes of scaling. Fourteen strategies, when paired with every other one, gave a total of 91 pairs $[\frac{(14)(13)}{2}]$. In the American English and foreign student groups each subject could be expected to judge only about 30 pairs due to time constraints. Therefore, the 60 native speakers of American English each judged about a third of the 91 pairs, providing a total 20 judgments on each pair of sentences. The 75 non-native speakers, judging about a third of the pairs, made the total of judgments for their group 25 for each pair of sentences.

Since more time was available for the Puerto Rican speakers, each of the ten subjects made a decision on all of the 91 pairs of sentences. The results of this method provided a rank order of the strategies from least to most deferential and a measure of psychological distance between successive strategies.

Results and Discussion. In the case of all three groups the raw data (i.e. the subjects' perception of politeness of the generic forms) provided a frequency matrix of the number of times each request strategy was judged more polite than every other strategy. This 14 × 14 matrix of frequencies was converted to a proportion matrix by dividing the cell entries of the frequency matrix by the number of judgments in each cell. Edwards' (1957) method then calls for a transformation of these proportions to a matrix of z scores. The z scores provide the normal deviate transformations for the proportions. Then, by summing and averaging the z scores associated with each request strategy a scale value of politeness is arrived at. This scaled score represents a quantity of politeness attributable to a particular request strategy by a given group of subjects.

The first hypothesis that native-speaking males and females would differ in their perception of the politeness of the fourteen request strategies was rejected. In fact, a Pearson product-moment correlation between male and female scaled scores was .90, significant at the .001 level. An examination of the rank orders of the generic forms shows the most disagreement on the relative politeness of *will* (strategy 1 in Table 1) and *you have* (strategy 8 in Table 1). In both cases the forms were perceived as more polite by males than females. In general, however, the agreement between male and female native speakers far outweighed the disagreement.

The correlation which corroborates the overall agreement between males and females in the perception of politeness masks some interesting differences, however. The graphs in Figure one (for males) and two (for females) may help to clarify some of those differences. Despite the fact that the request strategies are ordered in roughly the same way on both graphs, there are obvious differences in the magnitude of the scaled scores (i.e. the height of the bars) and the distances between successive scores (i.e. the distances between the bars). Males exhibit much smaller scores than females as well as a much narrower range from the most to least polite request strategy. These phenomena indicate two things: First, as a group, females display more unanimity about politeness than males do. Second, the large increments on the female scale, especially between strategies 11 (do) and 10 (can) and strategies 2 (you can) and 5 (I want), indicate that they are more categorical in their perception of politeness than males. Either a form is extremely polite or extremely impolite for a female speaker of English. Males, on the other hand, exhibit a less categorical perception of the fourteen strategies.

The perception of politeness among non-native speakers as compared with native speakers shows similarity to the congruence betwen male and female native speakers. The correlation between native speaker and non-native speaker perceptions was .53, again significant at the .001 level. Only two strategies (if and can) were outstanding in their differential ranks by native and non-native groups. *If you can give me some rice?* while ranked first by native speakers was ranked sixth by non-natives. Can, a fairly neutral strategy for natives (ranked seventh) was perceived as relatively polite by non-natives (ranked fourth). Several possible explanations can be suggested for these differences. Perhaps non-native speakers judged the sentence If . . . on the basis of grammaticality rather than politeness, causing it to achieve a lower score than for native speakers. Alternatively, native speakers may have provided additional linguistic content for this form, choosing *I wonder if* . . . instead of If. . . . Can might be grouped with other modals by non-native speakers in a category of relative politeness due to the polite nature of modals in their respective first languages. Can, since it is used the most by native speakers, takes on the character of an all-purpose request strategy. This heavy use may be perceived as politeness by somewhat naive non-native speakers.

The surface similarity between the native and non-native groups on the one hand, and male and female English speakers on the other, continues at a deeper level of examination. The graphs in Figures 3 and 4 depict the scale values for the 14 request strategies (in order from most to least polite) for native and non-native speakers. Comparing those graphs with the scaled scores for males in Figure 1 and females in Figure 2, we see the following: native speakers (Figure 3) look very much like males in both the magnitude of the scores (height) and the distance between successive forms. Non-native speakers, in contrast, exhibit much larger scale values attesting to the greater unanimity of their judgments. This phenomenon was especially true at the extremities of the politeness scale. This result implies that non-native speakers of English, like female native speakers, show a lot of uniformity in their perception of

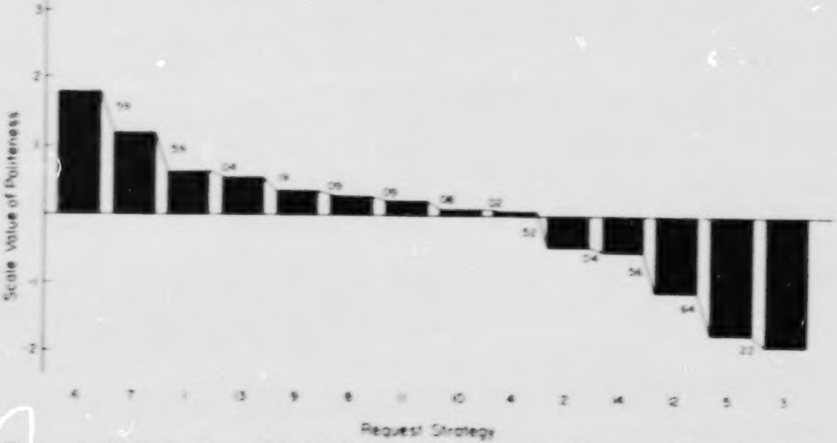


Figure 1. Changes in an individual student's writing speed (in words per minute) over 12 compositions.

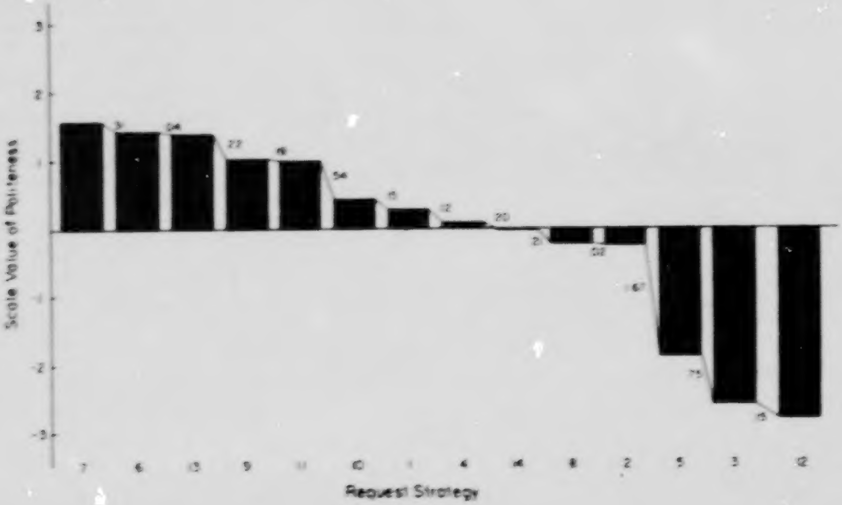


Figure 2. Changes in an individual student's grammatical error frequency (errors per 100 words) over 12 compositions.

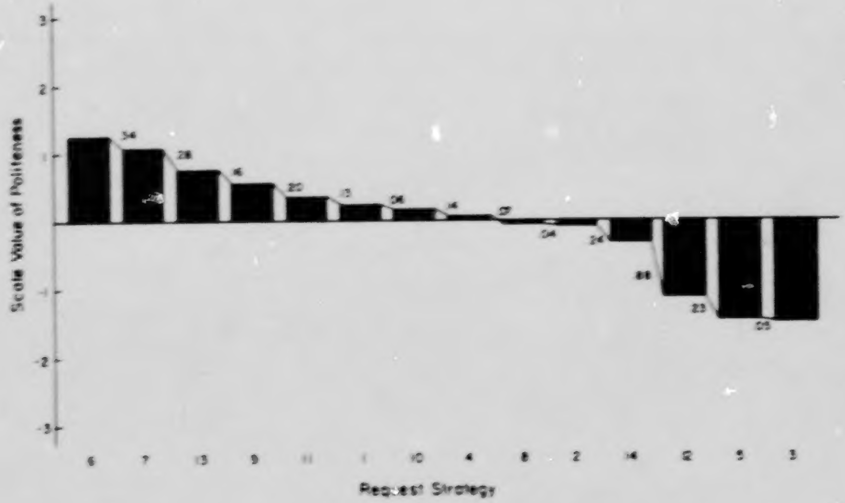


Figure 3. Changes in an individual student's misspelling frequency (misspellings per 100 words) over 12 compositions.

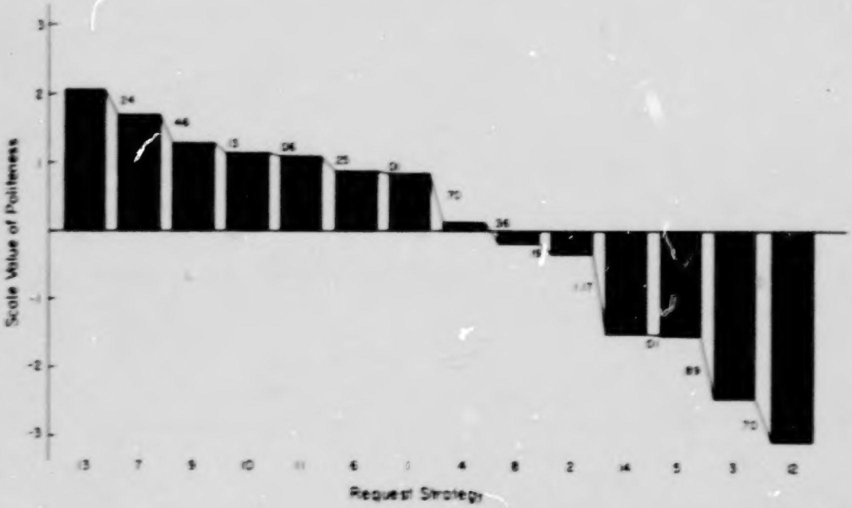


Figure 4. Graphic display of scale values of politeness for 14 English request strategies judged by non-native speakers (n = 25).

those strategies they judge to be extremely polite (may, could) and those they judge to be extremely impolite (you have to, imperative).

It is much easier to speculate about the nature of this finding in non-native speakers than in female native speakers. The greater overall agreement among foreign students than native speakers indicates less differentiation in the politeness of request strategies. These second language learners did not demonstrate the same gradual increase in politeness from strategy to strategy that native speakers displayed. Rather, the request strategies tended to cluster more for this group, especially at the impolite end of the scale. Thus, the psychological distance between the fourth and fifth least polite forms (Where and you can) represent almost 25 percent of the entire range of the scale.

One possible explanation for the clustering of the modals *may*, *would*, *could*, and *can* at the upper end of the non-native speakers' politeness scale is that these forms may have been taught in their English classes. The association of forms taught in the classroom with a politeness range seems most plausible. The lack of differentiation in the perceptions of the non-native group may stem from incomplete knowledge of the force behind the request strategies. The lack of exposure to as many varied contexts as native speakers have been exposed to also contributes to the clustering of strategies at the extremities of the scale of politeness.

In an effort to better understand the complex differences between native and non-native speakers' perception of politeness in English, we conducted a similar experiment in Spanish with Puerto Rican speakers. We expected the

Puerto Ricans to demonstrate the same high correlation between males and females that the other two groups display. The amount of contact between Puerto Ricans and North Americans, the similarity of request strategies in English and Spanish, and the amount of language contact all led to this expectation.

Although the range for the two scales is roughly the same (3.09 for males and 3.35 for females) the correlation between the politeness scores is only .52 which is not significant. Thus, unlike the high correlations in the previous two groups, one cannot predict female perception of politeness from male perception or vice versa in the Puerto Rican group. These sex differences are even more commanding when one considers the nature of the scales. Neither the male nor female scales exhibit any strategies with an extreme scale value of politeness. This means that there is a lack of unanimity among the judges, the same phenomenon which occurred in the native English speakers' ratings but which did not result in the non-native group's perception of politeness in English. The fact that the scales exhibited the same range but produced different results adds to the validity of the claim about sex differences in Puerto Ricans' perception of politeness, despite the small size of the sample.

Reviewing the data, a number of interesting conclusions and several pressing questions emerge. The data tell us the following: (1) Among native English speakers, males and females perceive the same request strategies as polite and impolite. Females, however, reflect a greater degree of unanimity and greater distinctions between polite and impolite forms than males do; (2) Native and non-native speakers of English show a very high correlation in their perception of the politeness of request strategies. Non-native speakers exhibit the same phenomenon with respect to native speakers as females do as compared with males. That is, the non-native speakers tend to agree more with each other and make greater distinctions between polite and impolite forms; (3) Puerto Rican males and females show no significant correlation in their perception of the politeness contained in Spanish request strategies.

We can conclude from this research that there may be profound cultural/linguistic differences in the perception of politeness (as evidenced by the high male-female correlation in the native English group and the lack of such a correlation in the Puerto Rican group). Furthermore, ESL instruction or something else in the second language acquisition process allows advanced non-native speakers to perceive the politeness of request strategies very much like native speakers. Their perceptions tend to reflect female native speakers' perceptions of politeness to a greater extent than they do male perceptions of politeness. Some of the questions arising from this work are: whether a developmental sequence exists in the perception of politeness; whether other linguistic groups will reflect the same sex differences that Puerto Ricans did in this study; and how the perception of politeness is mirrored in the production of these forms by adult second language learners.

Indirect Speech Acts in ESL: Indirect Answers

Patricia L. Carrell[1]

Southern Illinois University

Acquisition of competence in indirect speech acts—the saying of one thing to "mean that but also to mean something else" (Searle 1975:60)—represents an important aspect of an ESL learner's overall communicative competence. Indirect answers to questions are an interesting type of indirect speech act. Rather than answer questions directly, native speakers commonly respond indirectly. For example, a response to "Did you scrub the floor?" might be "Well, I swept it." Understanding such an indirect answer requires that the ESL listener have the ability to recognize and correctly interpret the communicative intent of the speaker when that intent is not directly, formally conveyed.

This paper reports an empirical study designed to test the competence of intermediate/advanced adult ESL learners to interpret the communicative intent of indirect answers. The performance of the ESL subjects is compared to that of a control group of adult native speakers.

Results indicate a significant difference between the two groups in interpreting the communicative intent of indirect answers. The ESL group did not perform as well as the native group. However, results also show a high level of ability among the ESL group, nonetheless. Implications of these findings for the preparation of communicative teaching materials are discussed.

1. Introduction

There has been considerable emphasis recently within EFL/ESL on the goal of preparing second language learners who are communicatively competent. As evidence of this emphasis one need only look in the meeting handbooks of the 1978 and 1979 TESOL conventions at the number of workshops and papers concerned with various approaches toward developing communicative or functional competence. Of particular prominence are the European (mainly British) approaches to development of functional/notional syllabuses and the related work in discourse analysis and English for specific purposes (Allen & Widdowson 1978, Candlin 1976). Taking a view of language as communication and a view of the understanding and producing of utterances as a pragmatic achievement, these approaches attempt to interrelate syntactic form and semantic interpretation with pragmatic use. They have a base in speech act theory—from the work of Austin (1962), Searle (1969) and others—and draw upon theoretical work in a wide range of other related areas, for example, studies in textual cohesion, in language function, in sociolinguistic variation, in presuppositional semantics, in ethnomethodology, in the ethnography of speaking, and others. (See Candlin 1976: 238-239) These approaches have the goal of offering to the learner "the language he will need to participate as speaker-hearer in real acts of communication." (Silva 1975: 342)

---

[1] The author wishes to thank SIU-C's Office of Research Development and Administration, which provided financial assistance for this study, including support for a research assistant, Gabriela Richter, whose help is also acknowledged. Thanks are also due to those on campus who granted the author access to the subjects used in this study, including Prof. Richard Lanigan of the Department of Speech Communication, and the Center for English as a Second Language (CESL).

I concur that this goal is what we should strive for in L2 teaching and learning, and I am in complete agreement with Candlin when he says:

> . . . the only proper goal of a language learning syllabus (is) that of leading a learner to be able to communicate and understand in a foreign language not only the meaning contained within grammatical structure, but also the range of meaning which lies outside the surface form, meaning as the communication of information which is negotiated between speakers and hearers in the context of their talk and against a backdrop of their beliefs. (Candlin 1976:238)

Since native English speakers are notorious for exploiting indirect speech acts, acquisition of competence in indirect speech acts represents an important aspect of an ESL learner's overall communicative competence.

However, I am interested in logically prior questions; namely, what are the communicative or pragmatic abilities of ESL learners? What are the communicative abilities of those who are learning ESL without the aid of specifically designed functional/notional syllabuses? In particular, how well do ESL learners interpret the communicative intent of indirect speech acts? The need for empirical studies to answer these basic questions is clear. Several recent papers by myself and others have attempted to give partial answers to these questions. One study (Flick & Carrell 1978) examined ESL learners' competence with indirect requests; in two other studies I looked at their competence with presupposed and asserted information (Carrell 1977) and with given versus new information (Carrell 1978). The present study attempts to shed further light on the communicative/pragmatic abilities of ESL learners, namely on their abilities to interpret indirect answers to questions.

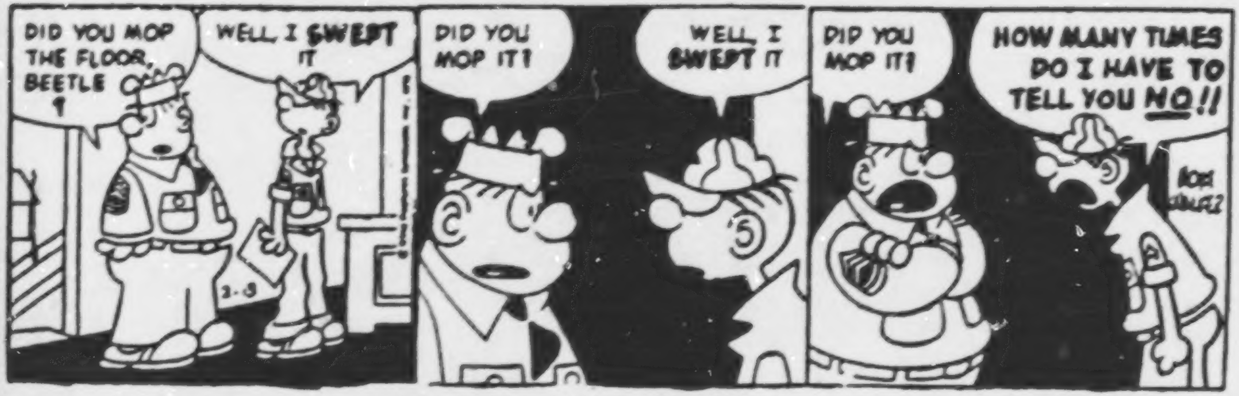
Indirect answers represent an interesting area of such empirical pragmatic investigation. Rather than answer questions directly, native speakers commonly answer indirectly. For example, an answer to "Did you scrub the floor?" might be "Well, I swept it."[2] Such an indirect answer is appropriate, informative and coherent. In order to understand the meaning of such an indirect answer, the ESL listener must have the ability to recognize and correctly interpret the communicative intent of the speaker when that intent is not conveyed directly or formally. In order to communicate competently, ESL learners must possess such abilities. The empirical question is, do they? And to what extent? This paper reports the results of a study designed to test the competence of intermediate and advanced adult ESL learners to correctly interpret the communicative intent of indirect answers.

As far as I know, as a category of indirect speech acts, indirect answers

---

[2] This example came to my attention in a recent Beetle Bailey cartoon strip, reproduced here with the permission of King Features Syndicate.

BEETLE BAILEY


DID YOU MOP THE FLOOR, BEETLE?
WELL, I SWEPT IT
2-8
DID YOU MOP IT?
WELL, I SWEPT IT
DID YOU MOP IT?
HOW MANY TIMES DO I HAVE TO TELL YOU NO!!

have not been studied in ESL. They have been the target of investigation in English as a first language by researchers in artificial intelligence and information retrieval systems. In a recent paper entitled "Why Ask?" Jerry Hobbs (a computer scientist) and Jane Robinson (a research linguist in artificial intelligence) analyze three categories of indirect answers with the aim of demonstrating that appropriate answers must address the higher communicative goals of the question and that various kinds of knowledge must be taken into account in analyzing the answer's appropriateness. (Hobbs & Robinson 1978)

2. Method

Subjects. The ESL subjects who participated in this study were 72 intermediate and advanced adult students at the Center for English as a Second Language at SIU-C. They came from levels 3, 4, and 5 of our five-level program. They were a heterogeneous group, but came predominantly from the Middle East, South America and the Far East. In the group there were 28 native Farsi speakers, 17 native Spanish speakers, 10 native Arabic speakers, 8 native Japanese speakers, and 9 speakers of other languages. CESL students placed in levels 3, 4, and 5 characteristically have Michigan and CELT scores averaging in the mid 50's, mid 60's and mid 70's at each of those levels, respectively. The curriculum for levels 3 and 4 consists of six hours per day, distributed equally among the basic skill areas: grammar, listening, speaking, reading, writing and language laboratory. Level 5 students attend only 2 hours per day during which time all skill areas are combined. The materials used at CESL are among the standards for such intensive centers in the U.S.

A control group of native adult speakers of English was used to insure that the performance of the ESL group could be compared to real performance data of natives, not merely to idealized, theoretically predicted behavior. The control group consisted of 29 SIU-C students enrolled in an undergraduate general studies course.

Procedures. The questionnaire consisted of 27 short, written dialogues. Each dialogue consisted of a contextualizing or stage-setting statement describing the conversational setting between two interlocutors. This was followed by one of the interlocutors asking a yes-no question. The question was then followed by an answer given by the second interlocutor. The subject's task was to draw a conclusion, in the form of a multiple choice response, about the meaning of the answer. In other words, the subject had to comprehend the information conveyed by the answer and come to a conclusion based on that information.

The design included not only indirect answers, but direct and unrelated/uncooperative answers as well. A direct answer was defined as one wherein the conclusion one comes to based on the answer follows strictly logically from the answer, or, in other words, is logically entailed by the answer; no other conclusion is logically possible. An indirect answer was defined as one wherein the conclusion, while not following logically from the answer, is a natural, likely conclusion to come to. It follows by a process of natural conversational inferencing (Grice 1975) or natural bridging (Clark & Haviland 1977). An unrelated answer was defined as one wherein the answer is not related to the question in any natural or obvious way, thereby blocking any natural bridging or inferencing about the conclusion. The design was further equally balanced among response choices, where affirmative, negative, or indeterminate conclusions were the expected responses. In the case of unrelated answers, only indeterminate conclusions were possible.

Figure 1. Research Design and Sample Items

DESIGN:		Answer-type			
Conclusion		Direct -a	Indirect -b	Unrelated -c	Total
	a-Affirmative	3 a-a	3 a-b	9 c-c	27
	b-Negative	3 b-a	3 b-b		
	c-Indeterminate	3 c-a	3 c-b		
	Total	9	9	9	

SOME SAMPLE ITEMS:

1. b-b Bob comes up to Ann in the Student Center. Bob says: "Did you go to the movies last night?" Ann says: "I had to study last night."
   a. Ann went to the movies last night.
   b. Ann did not go to the movies last night.
   c. I have no idea at all whether (a) or (b).
2. a-a Sue and Lisa meet in the shopping center. Sue says: "Did you buy a new dress?" Lisa says: "I bought one at Sears."
3. b-a Jim and Henry are talking about a party they went to the night before. Jim says: "Did you drink whiskey?" Henry says: "All I drank was beer."
4. c-a Cynthia and Maggie are talking about their ex-classmate, Lucy. Cynthia says: "Did Lucy speak French?" Maggie says: "I don't know."
5. a-b Sam and Dick are talking. Sam says: "Did you go to get a pizza last night?" Dick says: "I had nothing better to do."
6. c-b Kenney and Pete are talking about a mutual friend. Kenney says: "Was Joe sick last week?" Pete says: "I didn't see him last week."
7. c-c James and Harold are discussing Harold's grandmother in Texas. James says: "Did you write her a letter?" Harold says: "I received a letter from an old friend of mine last week."

A table of random numbers was used to randomly order the 27 items on the questionnaire.

The instrument was administered to the subjects in their classroom groups. Instructions were given both orally and in writing and example items were illustrated. Subjects were told to take as long as they required. In order to be certain that none of the vocabulary items posed any difficulty for the ESL subjects, they were told that they could raise their hands and ask about the meaning of any words they were unsure of.

Comparing the performance of the control group of native speakers to that idealized, theoretically predicted behavior, i.e. to a perfect performance, shows that the native group did not perform in the idealized manner. That is,

3. Results

The descriptive statistical results are reported in table 1.

TABLE 1

Descriptive Statistics

	Direct Answers	Indirect Answers	Unrelated Answers	Total
ESL N = 72	M = 8.43 SD = .802 R = 6-9	M = 6.15 SD = 1.507 R = 3-9	M = 6.97 SD = 2.[illegible]23 R = 0-9	M = 21.56 SD = 3.058 R = 11-27
Native N = 29	M = 8.55 SD = .854 R = 6-9	M = 7.52 SD = 1.163 R = 5-9	M = 7.28 SD = 2.347 R = 2-9	M = 23.35 SD = 3.272 R = 14-27
Perfect	M = 9	M = 9	M = 9	M = 27
Random	M = 3	M = 3	M = 3	M = 9

M (mean)
SD (standard deviation)
R (range)

there was a statistically significant difference between the native speakers' mean scores and perfect scores on each of the three answer-types and on their total performance as well. See table 2.

TABLE 2

Native Group Mean Scores Compared to Perfect Scores
One Sample t-tests

	Direct Answers	Indirect Answers	Unrelated Answers	Total
Native N = 29	8.55	7.52	7.28	23.35
Perfect	9.00	9.00	9.00	27.00
t-value	2.837	6.852	3.947	6.015
significance	$p < .01$	$p < .01$	$p < .01$	$p < .01$

This finding shows that in order to be fair to the ESL group, one ought not to compare their behavior to the idealized, theoretically predicted behavior but to the actual behavior of the native group.

On the other hand, comparing the performance of the control group to a totally random performance indicates that that group was not behaving totally randomly, as can be seen in table 3. In fact, the same may be said of the ESL group as well; see table 4. Neither group behaved in a random fashion, indicating that they were each attending to the tasks, attempting to perform them

TABLE 3

Native Group Mean Scores Compared to Random Scores
One Sample t-tests

	Direct Answers	Indirect Answers	Unrelated Answers	Total
Native N = 29	8.55	7.52	7.28	23.35
Random	3.00	3.00	3.00	9.00
t-value	34.994	20.926	9.821	23.609
significance	$p < .001$	$p < .001$	$p < .001$	$p < .001$

TABLE 4

Native Group Mean Scores Compared to Random Scores
One Sample t-tests

	Direct Answers	Indirect Answers	Unrelated Answers	Total
ESL N = 72	8.43	6.15	6.97	21.56
Random	3.00	3.00	3.00	9.00
t-value	57.460	17.736	15.867	34.839
significance	$p < .001$	$p < .001$	$p < .001$	$p < .001$

as well as possible, and performing them better than would have been expected if they'd been merely guessing randomly at the choices.

Incidentally, Cronbach-Alpha indices of the reliability of the questionnaire for each of the two groups, if the 27 items are projected to tests of 100 items, were .86 for the ESL group and .92 for the native group. These figures indicate that, although the questionnaire was far from perfect, and by no means a normed, standardized instrument, it was a fairly reliable measure.

Pearson product moment correlations among the sub-tests and their respective significances are reported in the correlation matrix in table 5 by groups; the native control group's correlations are above the diagonal, the ESL group's correlations are below the diagonal. For both groups, the only significant correlations are between direct answers and unrelated answers. For the native group $r = .46$, $p = .006$; for the ESL group $r = .40$, $p = .001$. The correlations involving indirect answers, both with direct answers and with unrelated answers, are not significant. This may indicate that the inferencing or bridging required in interpreting indirect answers is a different ability, different in kind, from either

TABLE 5

Sub-test Correlation Matrix by Groups
Pearson Product Moment Correlations and Significances
Native Group above diagonal, ESL Group below diagonal

	Direct Answers	Indirect Answers	Unrelated Answers
Direct Answers	—	.20 p=.15	.46 p=.006
Indirect Answers	−.02 p=.43	—	.16 p=.20
Unrelated Answers	.40 p=.001	.10 p=.21	—

interpreting direct answers or determining that an answer is unrelated to the question and thus coming to an indeterminate conclusion.

Within the ESL group, although there were three instructional-proficiency levels represented (30 subjects at level 3, 31 subjects at level 4, and 11 subjects at level 5), there were no statistically significant differences among the three levels in their performance on the three answer-types.[3] Nor were there any statistically significant differences among the five native language groups represented (Farsi, Spanish, Arabic, Japanese, and other).[4] Therefore, for the purposes of this study, all ESL subjects were treated as members of a single group, differences in their English proficiency and native language being statistically irrelevant to the basic research question of this study.

Thus, the basic comparisons of interest for the present study are those between the ESL group taken as a whole and the native control group. First, comparing the performances of the two groups on the three answer-types, we see that the performance of the ESL group was slightly lower overall than that of the control group of native speakers but was nonetheless quite high. Compare the native mean scores of 23.35 total, 8.55 direct answers, 7.52 indirect answers, and 7.28 unrelated answers, to the ESL mean scores of 21.56 total, 8.43 direct answers, 6.15 indirect answers, and 6.97 unrelated answers, respectively, in table 1. As previously shown, these scores for both groups are much better than chance. The sub-test scores for both groups can be compared, both univariately and multivariately. If viewed univariately, as a series of t-test comparisons for independent samples, the only statistically significant difference between the groups is in the area of the indirect answers. See table 6.

---

[3] Comparisons among the three proficiency levels were made in two ways. First, viewed as a series of univariate (single dependent variable) comparisons, oneway analyses of variance were performed, comparing the three proficiency levels on each of the answer-types separately. However, since some of the sub-tests are modestly correlated, it was also decided to view the comparisons as a multivariate problem (multiple dependent variables), comparing the three proficiency levels on all three answer-types simultaneously. Viewed either way, there were no statistically significant differences among the three proficiency levels on the three answer-types.

[4] As above, the comparisons were made both univariately and multivariately. Viewed either way, there were no statistically significant differences among the five native language groups on the three answer-types.

TABLE 6
Comparisons of Native Group and ESL Group
on Three Answer-Types
Univariate t-tests for Independent Samples

	Direct Answers	Indirect Answers	Unrelated Answers
Native N=29	8.55	7.52	7.28
ESL N=72	8.43	6.15	6.97
t-value df=99	.67	4.36	.63
significance	p > .50	p < .0001	p > .53

If viewed multivariately, as a problem of distinguishing between the two groups on the three answer-types as multiple discriminating dependent variables, the same result is obtained. The significance of the Wilk's Lambda in table 7 indicates an overall difference between the two groups on the three answer-types. A discriminant analysis identifies the discriminant function—the linear combination of the dependent variables and their relative weights—which best discriminates between the two groups; the standardized discriminant function coefficients determine the relative contribution of each of the three dependent variables to the discriminant function. As can be seen in table 7, most, if not all, of the discrimination is due to one dependent variable—indirect answers. The relative sizes of the coefficients indicate that the discrimination between the two groups is not due to the interaction of the three dependent variables, but is primarily due to the one dependent variable—indirect answers.

In other words, the ESL group did fairly well on all parts of the interpretation task—ESL learners performed comparably, in fact, to the native group on direct

TABLE 7
Comparison of Native Group and ESL Group
on Three Answer-Types
Multivariate tests—Wilk's Lambda/Discriminant Analysis

Discriminant Function	Eigenvalue	Wilk's Lambda	Chi-Square	DF	Significance
1	.19473	.8370	17.347	3	p < .001

Standardized Discriminant Function Coefficients

Direct Answers	− 0.117
Indirect Answers	− 0.989
Unrelated Answers	+ 0.021

answers and on unrelated answers—but they did differ significantly from the control group of native speakers in performance on interpreting the communicative intent of the indirect answers.

4. Discussion

The results yielded basically two interesting findings. First, there was a significant difference in the performance on interpreting the communicative intent of indirect answers between the ESL group and the native group. The ESL group did not perform as well as the native group; they do not possess native-like competence in this area. Second, however, the ESL did quite well on interpreting all three answer-types, including interpreting the communicative intent of the indirect answers—far better than a chance performance. In other words, while the ESL subjects are not comparable to native speakers in this area of communicative competence, they exhibit a high level of ability nonetheless.

The item which gave the ESL subjects the greatest difficulty were two indirect answers in which the answer was itself a question—a question with a supposed obvious answer—an answer like "Is the Pope Catholic?" According to Jerry Morgan (1978), the schema for drawing an implicature in these cases is: Answer an obvious yes/no question by replying with another question whose answer is very obvious and the same as the answer you intend to convey." (Morgan 1978: 278) According to Morgan, most Americans have two or three stock expressions like this, but new ones may be manufactured as they are needed. In this study, one of these obvious answers was an obvious affirmative, the other an obvious negative:

Affirmative: Q: Is Becky rich?
A: Do birds fly?
Negative: Q: Is Greg smart?
A: Is Sadat Jewish?

Somehow the obviousness of the answers eluded many of the ESL subjects. Not surprisingly, on several grounds—both psycholinguistic and other, the negative proved even more difficult than the affirmative. While only 58% interpreted the affirmative item correctly, only 40% interpreted the negative item correctly (despite its timely relevance in the fall of 1978 when the test was administered). Interestingly, the control group also had its worst performance on interpreting the negative item—only 62% got it correct. (93% got the affirmative item correct.)

The generally fine performance by the ESL subjects on interpreting indirect answers indicates that they already have a high level of proficiency in comprehending these kinds of indirect speech acts. We might speculate that a large measure of this is carried over from first language acquisition or may, perhaps, reflect universal linguistic or communication properties. Although all of the research that I am aware of in the area of indirect speech acts has been done on English, it seems reasonable to suppose that all languages have mechanisms for performing indirect speech acts. Searle (1975) argues:

> . . . the mechanisms by which indirect speech acts are meant and understood are perfectly general—having to do with the theory of speech acts, the principles of cooperative conversation, and shared background information . . . (Searle 1975:75)

While there are obviously going to be differences among languages in the specific indirect speech forms used to perform different indirect speech acts, (e.g. "Can you hand me that book?" may function as an indirect request in English, but its Czech translation will sound odd if uttered as a request in Czech—Searle 1975: 76), we may speculate that classes of indirect speech acts exist universally—e.g. indirect requests and indirect answers. Thus, foreign students approaching the learning of English as a second language may, if only subconsciously, expect to encounter indirect speech act situations, communicative situations wherein, as Searle (1975) says:

> the speaker communicates to the hearer more than he actually says by way of relying on their mutually shared background information, both linguistic and nonlinguistic, together with the general powers of rationality and inference on the part of the hearer. (Searle 1975: 60-61)

5. Conclusions

The implications of these findings for the current emphasis in EFL/ESL in preparing functional/notional syllabuses, the so-called communicative approaches mentioned at the outset appear clear. If we are to construct the most effective of these syllabuses, utilizing and building on the underlying communicative abilities which are already present and not needlessly re-teaching something already acquired, we must understand more about these underlying communicative competencies. In other words, we need more studies like this one to get at existing communicative competence. We must understand more about the extent to which these abilities are present in order to establish the base-line data upon which to build. This study shows that in the area of interpreting indirect answers, there is much there to be built upon as syllabuses are written to aid ESL students to achieve native-like communicative competence.

Questions in Task-Oriented Discourse

Ruth Crymes and William Potter
University of Hawaii

An analysis of the substantive questions in four task-oriented discussions in English, each of which addressed the same topics but had different participants, revealed the following question functions: *leading* questions, *open* questions, *challenges*, *softened assertions*, and *clarification* questions. These question functions, except for clarification, are related to the nature and structure of task-oriented discourse, but the form of the questions in relation to function and the distribution of the questions in different parts of the discourse are related to the interactional patterns of each particular group of discussants. The nature and structure of task-oriented discourse is described and the interactional patterns of the four different groups of discussants are analyzed and compared.

We have investigated samples of one kind of oral discourse, the task-oriented discussion, focussing on the questions that were asked in order to see what connections we might find between the questions on the one hand and the structure of the discourse and the roles of the participants on the other. We looked at both question functions and question forms.

We have analyzed four discussions in English, two by native speakers, one by non-native speakers, and one by a mixed group of native and non-native speakers. We found that the three non-mixed groups—that is, the groups that were either all native or all non-native speakers—each executed a similar discourse plan but handled its elements and exploited the resources of the language in different though systematically related ways. Each of these groups succeeded as well as the others in communicating within the group. In contrast, the mixed group of native and non-native speakers had considerable difficulty in carrying out the discussion and communicating with each other.

We will first describe our procedure for collecting the data. Then, in presenting the data and our analysis of it, we will discuss (1) the structure of task-oriented discussions, (2) the roles of the participants, (3) the functions of the questions which appeared in the discussions as they relate to the structure of the discourse and to the roles of the participants, and (4) the forms of the questions as related to their functions. Finally, we will compare the four discussions in terms of these four parameters and conclude with some implications for teaching.

1. Data Collection

The four task-oriented discussions which we audio-taped and transcribed, following the system of Sacks, Schegloff, and Jefferson (1974), were on ESL topics. The discussants were ESL professionals, the non-native speakers among them all having a high degree of proficiency in English.

The four discussion groups had the following compositions: (1) four native speakers of English, Americans in mid-career—three men and one woman; (2) four native speakers of English, Americans at the beginning of their careers, all candidates for the MA degree in ESL—two men and two women; (3) four non-native speakers of English in mid-career from Asian and Pacific countries—two men and two women from Burma, Fiji, Korea, and Thailand; and (4) a mixed group of native and non-native speakers of English in mid-career—two men and two women from Japan, New Zealand, the Philippines, and the United States. The discussions were 12, 26, 28, and 30 minutes long, respectively.

The procedure was to elicit a discussion by first giving a written questionnaire to each of the participants which was patterned after a values clarification exercise (Simon, Howe, and Kirchenbaum 1972: 252-254), asking them to strongly agree, somewhat agree, somewhat disagree, or strongly disagree with the following five teaching practices: Teachers should (1) expect students to read aloud with good pronunciation; (2) emphasize vocabulary acquisition from the beginning; (3) make a lot of use of dialogues; (4) give grammatical explanations to their students; and (5) teach simple structures before complex ones. The participants were then asked to compare their answers and in each case where there was not unanimity they were to discuss the statement. As it happened, there was no unanimity in any group on any statement.

Then the task was for the participants to come to a consensus on each of the issues raised by the statements, revising the statements as necessary. They were instructed to do this orally. We told them that we would listen to the tape later and summarize their ideas. We did not tell them that our purpose was to analyze questions in the discourse; we told them instead that we wanted to compare the ideas of the different groups on ESL pedagogical practices. Topics from any area could have been used, but we had access to ESL professionals, and so we solicited their cooperation.

2. Structure of the task-oriented discussion

A task-oriented discussion is one in which participants have not only an ideational focus for the content of their utterances but also a charge to try to reach some community of thought about that focus which has the potential of being coherently summarized. To reach a community of thought does not require agreement; it requires only an effort to understand the whole picture as it emerges in the discussion, an effort which might then set the stage for points of agreement or for the decision to agree to disagree or simply for a summary. A task-oriented discussion, in its overall function, may be an oral, experiential equivalent of the review of the literature in a research report. Participants are set, or set themselves, to the task of determining and/or characterizing the views of their group on some topic or issue.

Preferably there would be more than two participants in such a discussion since the goal is to ascertain and benefit from collective knowledge and experience, but it would be possible to have two participants only. The upper limits are flexible within the contraints that space and time impose on oral interaction, but the discussion would shift from informal to formal as the numbers got larger.

The most ritualized form of the task-oriented discussion is the panel discussion. Less ritualized forms occur, for example, in staff and organization meetings, in classrooms or conferences where small groups are collectively addressing an issue, and in committees charged with preparing position papers or framing resolutions.

In looking for the structure of the informal task-oriented discussions that we taped, we referred to the ritualized form, the panel discussion, and noted that it has a beginning, a middle, and an end. At the beginning is the presentation of each panelist's view of a set topic; in the middle, opportunity for open discussion (which may or may not be seized); and at the end summation by a moderator, who has also presided throughout.

This structure is also visible in informal task-oriented discussions. There is a presider and moderator, roles which we will describe below. And each topic is discussed in a sequence of three parts which we have labeled the *presentation*, the *exploration*, and the *accommodation*.

During the presentation each participant explains his or her position relative to the topic. The presentation ends when all participants have done so. During the exploration participants expand on what they presented earlier, respond to the ideas of others, and seek clarification as they feel necessary. When they have said what they have to say within their allotted time, they try to come to some kind of accommodation. They may take a common position, they may explain why they cannot take a common position, or they may just give up and stop discussion of that topic. Or they may, during their efforts at accommodation, especially if they feel that someone is pushing accommodation prematurely, re-open the topic for further exploration and then later move into the accommodation stage again.

3. The roles of the participants

We found that in the discussions the participants could assume three different roles. All assumed the role of *discussant* by contributing to the substance of the discussion. In addition, one or more than one self-selected the roles of *traffic controller* and *middleman*.

The traffic controller presided, managing procedural matters. He or she got the discussion going and tried to get it and keep it on track and conclude it. One participant might retain this role throughout a discussion, participants might take turns, or they might even at some times assume the role jointly. In our data traffic controllers said such things as the following:

(1) Non-native speakers: O: OK then. We'll go down the list then. It's

about-//O: Uh, OK. Carry on with your, reasons.//O: Yeh, so the point is, shall we se- settle on A[gree] S[omewhat] or shall we settle on D[isagree] S[omewhat] (sigh). S[trongly] A[gree] and S[trongly] D[isagree] are out of the question now.

(2) Native speakers/mid-career: C: Number one. I'm uhs, uh, strong disagreer.//C: How about you?//C: Shall we, can we go to, compromise on disagree somewhat?//C: You could live with that?

(3) Native speakers/beginning career: B: Uh (2 sec) for the- questions first of all how did you interpret them? For what level? Kay did you lump them all together under one level er, when you were, thinking of these items?//T: So I don't know, if the're that many, = B: That we can make— that you could— T. = general statements that we can make. B: Maybe we can put in qualifiers.// M. What about four "Give grammatical explanations," = B. "to the students." T: What did you—

The middleman mediated, managing substantive matters. This role, too, might be assumed by one person throughout, who might or might not also be the traffice controller, or the role might shift from one participant to another. The middleman tried to help the others formulate their ideas, clarify what others had said, and make generalizations based on what others had said. For example, note the utterances of G (native speaker/mid-career) and O (non-native speaker) in the following excerpts.

Topic: Teachers should emphasize vocabulary acquisition from the beginning.

W. I said "Disagree Somewhat." Uh (2 sec) Mm. (5 sec) Well, I don't know, I, hm, I feel that tuh they'll learn, what they need to learn. Now- "from the beginning" I don't know that's the hard part but, (3 sec)

(4) G. (Well) would you put it in their way? Would you put, new vocabulary in their *way*, so they'd *have* to learn? (2 sec)

W. No I wouldn't.

(5) G. You wouldn't, deliberately (do anything about it)?

W. No, of course I'm envisioning the idea of—the ideal classroom, where the students interact and learn English, they'll learn, that vocabulary right? (Laughs)

Topic: Teachers should have students read aloud with good pronunciation. (near end of discussion)

(6) G. Yeh. Well it seems two things are mixed up. You know, if you're teaching uh pronunciation, you're uh probably not at the point where you're going to have them read aloud anyway. It seems as though the two things are, out of focus.

Topic: Teachers should give grammatical explanations to their students.

(7) O. What would be your opinion, of giving grammatical explanations
at the, initial stages [of teaching?]
L. [No] not at all
I don't [think that.]
(8) O. [You would] strongly disagree then?
L. No, I would strongly disagree.
(9) O. So you're you're speaking in terms of the level of your [L. Right.]
learners. Yeh?
L. I guess all my answers here [because I have=
(10) O. [Yes, of course ( )
yeh related to your-]
L. =decided to- base my answers on-]

O then turns to Q and asks her to give her reasons.

A person asking a leading question, which O's and G's questions (4, 5, 7) above exemplify and which we will discuss below, was by definition assuming the role of middleman. A middleman also makes mediating and summarizing statements as in (6), or even statements for another person as in (8, 9, 10).

In asking questions, the participant in the role of middleman may sometimes be trying to get a particular answer, especially in the accommodation stage, since, unlike the moderator in a panel discussion, who is only a middleman, the middleman in these informal task-oriented discussions also wears the hat of discussant, as illustrated below in (11, 12).

Topic: Teachers should teach simple structures before complex ones.

G. Suppose you take a (task) of learning for instance.
C. Yeh. OK.
(11) G. Do you teach the simple before you teach the complex?
C. Yeh. (3 sec)
C. (To W) Then you gotta *move* then.
(12) G. Is it any different with language? (4 sec)
W. °Maybe.
R. Might be, yeh. Might be that you teach the complex and then you find out what the simple things are that make up the complex. I don't know. (4 sec)
G. Well I'd be willin to move.
W. (Laughs)
G. I'm not sure=—uh what I meanhhhh.
(R and G laugh. End of discussion on this topic.)

G has tried to lead others to his position. But he ends up uncertain about what his own position is.

4. Question functions

The questions that we examined were those which addressed substantive matters relating to the topic under discussion. We did not examine rhetorical questions, questions about procedural matters, or questions not relevant to the topic. In identifying which utterances were questions we used as criteria both the various question forms and the expectation of a response.

We identified five question functions as follows: *open* questions, *leading* questions, *challenges*, *softened assertions*, and questions of *clarification*, this last of two types—one asking for confirmation of the acoustical signal and the other of meaning.

We believe that these five categories of functions are exhaustive for our data, but we recognize our whole framework as preliminary and we call attention to the ultimate subjectivity of our interpretations of speaker intentions. Following are discussion and illustration of each of these question functions.

Open questions seek information. They are for the benefit of the questioner. In the exploration stage of the native speaker/mid-career discussion of "Teachers should have students read aloud with good pronunciation," W and C make it clear from their responses to G and R's reasons that they don't think that good pronunciation has any relevance to reading. But the four discussants don't seem to understand that W and C are associating the reference to "reading aloud" with the reading process and G and R are associating it with oral production. G asks an open question of W and C.

(13) G. Well if you two had them read aloud, in the first place, what would you expect?
C. Oh I'd try for uh, well I don't know if I would uh, I would have em read aloud.
W. °Yeh, but uh, you wouldn't expect them to, [C. °Yeh.] read aloud.

G has tried to seek information, but C and W have answered evasively because the question is meaningless to them. They are operating on a different assumption from G about the purpose of reading aloud.

In the native speaker/beginning career discussion of "Teachers should teach simple structures before complex ones" we find the following open question in the exploration of what complex means.

(14) B. . . . °What else would be complex?
T. Well of course it's- all very relative to what you're teaching, I mean some of the things we're talking about, I suppose might be considered yaknow simple, to one class and very difficult to another, even, even tenses, something like past perfect, and present perfect, some people yaknow I've had students who have real trouble keeping the, the time structure right. You never know.
B. Yeh I would think those are more complex.

Leading questions help another formulate or clarify his or her ideas or try to lead the group to reach points of agreement. Whereas open questions are for the benefit of the questioner, leading questions are for the benefit of the respondent. In asking a leading question, a participant is assuming the role of middleman, as illustrated in (4, 5, 7) above, which occurred in the presentation stage, and in (11, 12) above, which occurred in the accommodation stage.

Challenges call into question the reasonableness of someone's assumption as the questioner perceives that assumption. In the native speaker/mid-career discussion of "Teachers should have students read aloud with good pronunciation" the excerpt following (13) above contains three challenges addressed by R to W as follows:

W. °Yeh, but uh, you wouldn't expect them to, [C. °Yeh.] read aloud.
(15) R. Never? Can't think of any circumstances under which you'd have, people read aloud?
W. Oh maybe uh, you know "What what sentence you got there" f- for [teaching writing.]
(16) R. [How about poetry?]
W. Oh no. I wouldn't uh,
(17) R. Poetry shouldn't be, said aloud? read aloud?
C. [Performed aloud]
W. [Mm.] The idea of good pronunciation is the part, particularly- (Laughs) (You'd say uh) correcting their pronunciation all the time.
R. I didn't say anything about correction. . . .

A challenge often provides more insight into the ideas of the questioner than the answer provides into the ideas of the respondent. Considerable exploration is often required to determine if challenger and respondent are talking out of the same assumptions or not, and they may never identify their source of difficulty.

The clearest example of such difficulty occurred in the group of two native and two non-native speakers. The topic was whether or not teachers should have their students read aloud with good pronunciation. All disagreed except Y (from Japan), who agreed somewhat. The discussion got shifted from exploring the relation between reading and pronunciation to the topic of what the standard of pronunciation should be. The stage was set for this shift when Y said that the situation in Japan was special, that the learners there needed to read aloud because the only way that they could practice speaking English was to have their teachers, who were not good English speakers themselves, "manipulate good materials," as he said, for choral reading for the purpose of providing a model for the students. What he apparently meant was that in Japan the teacher would play tapes, the learners would listen and then later read out loud what they had listened to and the teacher would evaluate the students' pronunciation. His assumption, as we interpret the total discussion, was that a teacher does not have to have good pronunciation to help the students acquire good pronunciation.

After A (from the United States) first ascertained that Y was supporting reading aloud as a pronunciation exercise, she then asked him an open question, "And the, who is the model for that?" When he answered that recordings were the model, A came up with three challenges in the ensuing discussion:

A. But if, if you're using the reading as a pronunciation exercise,
for your students in Japan and you said that the teacher, uh,
many times didn't have very good pronunciation
(18) [T. No:.] then who is the model, who do the tea- do the students,
model their pronunciation, after so that they get the good pronunciation, when they read?
Y. Ah: : They have the record, they have the tapes and the: : : they
have lots of materials for example TV program and radio program as well.
Z. Are you suggesting that there wouldn't be any transfer?
A. Well I, I think if you were going to *use* it in this *way*, I'm I
just- I do- I mean that, I wouldn't use it that wayhhh. But if you
we:re=
Y. Since we are *not* native [speakers,] we have to use,
A. [Yeh.]
A. Mmhmn. But what my poi-, my point is, if you're going to- use
it as a pronunciation exercise,
(19) how do they know whether they're pronouncing it, uh like
[uh,]=
Y. [You] mean the-
A. =the teacher wants it pronounced? If and especially if the
teacher, i- if he, as you said, if the teacher doesn't have good
(20) pronunciation, then how can they improve good pronunciation,
by reading, out loud?
Z. Do you think that this is an important facet F? This, pronunciation?. . . .

Apparently A concluded that there was no reasonable answer to her question, but she kept asking it in order to get her own point across, which, however, she failed to do. Y stuck to his idea that Japanese English teachers have to use recordings in a reading aloud lesson since they are not native speakers. Z's leading question to A, "Are you suggesting that there wouldn't be any transfer?" shows that Z has some notion of Y's assumption. Z thinks that the whole discussion is beside the point but evidently does not want to offend Y, and so he addresses a challenge, at the end of the above exchange, not to Y but to F, the other non-native speaker, from whom he expects and receives agreement: "Do you think this is an important facet F?" F agrees that it is not. Z is trying to get the group back to the original topic, but he does not succeed. They shift to talk about what the standard of pronunciation should be, Japanese English or native speaker English, and they never get back to the original issue.

Softened assertions state one's belief or understanding and at the same time seek confirmation of it from others. In the native speaker/mid-career discussion of teaching vocabulary from the beginning, G has tried to lead W to make her position clearer (see examples 4 and 5 above). He has asked: "You wouldn't deliberately (do anything about it [vocabulary]?" W's answer concludes with a softened assertion directed to the whole group.

W. No, of course I'm, en-envisioning the idea of- the ideal class-
room, where the students interact and learn English, and what-
(21) ever they needs they have, they'll learn, that vocabulary right?
(Laughs)
C. Wrong.
(All laugh)

When asked by a middleman, a softened assertion states the middleman's understanding of the group's position, as in the following excerpt from the native speaker/beginning career discussion of teaching vocabulary.

B. In other words these restricted books in the beginning where
they have- a limited vocabulary like if you have a beginning
class, [M. Mmhmn.] uh, the're several series that have very
limited [vocabularies.]
T. [Seven hundred [wor]ds, twelve=]
(N) [Guided vocabulary.]
T. =hundred words,
B. Yeh.
M. Hnh.
(22) B. And that ch- you-think that right from the beginning there
should be a beefing up of vocabulary?
M. Attention drawn to it. Uh I don't think you have to use con-
trolled materials.
B. Yeh, OK.

In our data questions of clarification occurred only in the mixed group discussion. This is probably partly happenstance and partly a consequence of the fact that when people assume communication is proceeding, as in the non-mixed discussions, they don't feel the need for asking questions of clarification so much.

At one point in the mixed group discussion, A needs to confirm what was said because of overlapping talk. She has asked Y whether he would use reading aloud as a reading exercise or as a pronunciation exercise.

Y. Mm. Yes. Choral reading-
A. As-
Y. [Group reading.]
F. [Pronunciation.]

(23) A. As pronunci[ation?]
Y. [Pronuncia]tion, yes.

It was common in this discussion for F (non-native speaker) to answer for Y (non-native speaker)

In the discussion of whether teachers should teach simple structures before complex ones, F presents her argument against structurally graded materials and Y asks a question to confirm his understanding of what kind of materials she is referring to in (24) below:

F. Well. Uh, if we uh (2 sec) in in structural linguistics, uh the audio-lingual approach, we are told or taught, that it's uh, °you know, it's pedagogically effective if we start with simple structures before we move on to complex ones, but uh, actually there is no, uh psychological reality to this kind of theory. . . . Now, uh, I think if you structure your syllabus in this manner, you are uh restricting as I said, the progress of the student and you come up with, with eh, lessons probably that is not eh, that is uh, that doesn't reflect reality because in reality students actually make use of a lot of embedded structures and this type of thing so- I don't think there is psychological reality to, uh to this uh kind of belief that y- you, yaknow, you succeed uh better if you start with the simple structures before you move on to, the more complex ones.
(N) Mm.
Y. I see.
Z. That's quite true, you know, that's borne out in our country as far as the teaching of reading is concerned.
(24) Y. Are you referring to some kind of the er automatical uh automatic, the substitution drill? or the parroting the
[uh changing] from one to another?
F. [Yes that's-]=
F. =Yes, tha- that's based on this.

5. The form of the questions

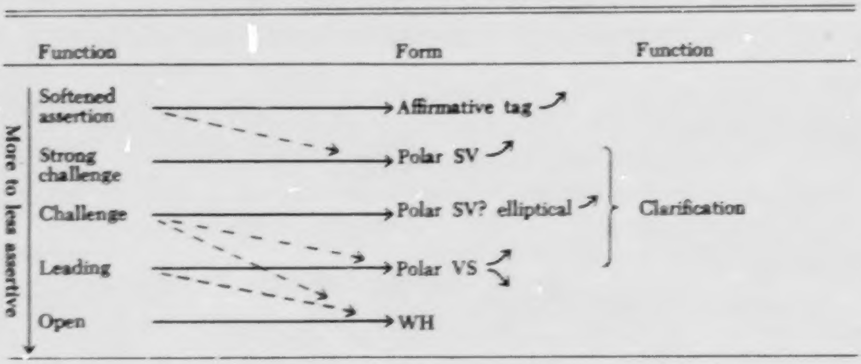
Elsewhere we have reported on the forms of the questions asked in the three non-mixed discussions (Crymes and Potter 1979). We noted that in the question forms there was a cline of assertiveness from most assertive to least assertive as follows: Affirmative Statement with Question Tag → Polar Subject Verb → Polar Subject Verb Elliptical → Polar Verb Subject → WH. For the question functions described in the preceding section, the unmarked forms, excluding questions of clarification, were as follows: Affirmative Statement with Question Tag for softened assertions, Polar Subject Verb for strong challenges, Polar Subject Verb elliptical for challenges, Polar Verb Subject for leading, and WH for open questions. But not only were there stronger and less strong challenges, there were also marked forms which were less assertive for the other functions, as follows: Polar Subject Verb instead of Affirmative Statement with Question Tag for softened assertions; Polar Verb Subject instead of Polar Verb Subject Elliptical for even weaker challenges; and WH instead of Polar Verb Subject for leading questions. Note, for example, (7) above, where the WH question is a leading question. It is answered with a "no" as if it were a Polar Verb Subject question. We called these marked forms downshifting, meaning that there was a shift from more to a less assertive form.

In the mixed group discussion we found examples of double downshifting. These were WH questions which were not interpretable as open questions. They were challenges. Evidently in these cases the questioner was challenging not what someone had said but the existence of an answer. That is, the questioner evidently felt that no reasonable answer was possible. But not wanting to come on with too strong a challenge, the questioner downshifted two steps to the WH form. Examples are (18, 19, 20) above.

As we have indicated, it was only in the mixed group discussion that we found questions of clarification. They were all polar questions of various forms. They were not challenges, because it was not clear what, if anything, could be challenged, nor leading questions, because they referred back to what had already been said. The questions of clarification in our data appear to be side-sequences (see Jefferson in Coulthard 1977:73-74).

See Chart 1 for a summary of the correlations between question functions and question forms.

TABLE 1
Question Form/Function Correlations
(Broken line = downshifting)



6. Summary and conclusions

Chart 2 provides characterizations of the four discussions.

The native speaker/beginning career group was the least managed and often built up communal utterances for both procedural and substantive communications. They asked more questions than the non-native speaker/mid-career group but fewer than the native speaker/mid-career group. Half their questions were open and a little under a third were challenges. Almost half were downshifted and almost all occurred during exploration stages.

The native speaker/mid-career discussion was quite firmly managed though there was some sharing of authority. The participants asked the most questions even though they spent much the shortest time in discussion, and all of them were about equally active as questioners. They asked almost half leading questions and a third challenges and distributed their questions about equally over the three stages. There was no downshifting.

The non-native/mid-career discussion was the most firmly managed, both procedurally and substantively, and the fewest questions were asked by the fewest participants. Half the questions were leading and over a third were challenges. Most occurred during the presentation stage, and somewhat more than half were downshifted.

The non-mixed groups thus all differed from each other in quantity and quality of management, number of questions and questioners, and distribution of the questions in the different parts of the discourse. But the native-speaker/beginning career and non-native speaker/mid-career speakers shared a propensity for downshifting, and the mid-career speakers, both native and non-native, favored leading questions and challenges whereas the native speaker/beginning career speakers favored open questions. Age, professional status, and cultural background were all operating in various combinations as influences.

In contrast, the mixed group had difficulties moving along in their discussions, both procedurally and substantively. There was a struggle for the role of traffic controller. The role of middleman was seldom assumed. The participants asked the most questions of any group, but all but one of the questions were asked by the two native speakers. Over half were questions of clarification and almost a third were challenges, most of the latter doubly downshifted. Further, it was only in the mixed group discussions that one person answered for another: one of the non-native speakers quite often answered for the other non-native speaker.

At least one or the other of the two of us investigating this data knew personally each of the discussants, and we recognize that some of the interaction that we have noted stemmed from individual personalities. Yet the three non-mixed discussions were different in degree rather than in kind from each other, and the mixed group was different in kind rather than degree from the other three.

Excluding the questions of clarification, the functions of the questions appeared to be primarily determined by the nature and structure of the discourse, and their form and distribution appeared to be primarily determined by the interactional pattern of the group. Where there was not a common interactional pattern, as there was not in the mixed group, there may not in fact have been any more lack of mutual understanding of the substance of the discussion than in the other groups, but there was more unease and frustration. The number of clarification questions asked were in part a manifestation of this frustration.

TABLE 2

Characterizations of Four Discussions

NS/BC = Native speakers/beginning career NN/MC = Non-native speakers/mid-career
NS/MC = Native speakers/mid-career MG = Mixed group/mid-career

	Roles of questioners		
	Traffic Controller (TC)		Middleman
	Initially	Thereafter	
NS/BC	Highest status (seniority)	Shared by all	Assumed by all as a communal persona
NS/MC	Highest status (rank)	Shared by two others but highest status predominates	Primarily the one who is never TC, minimal participation by chief TC
NN/MC	Oldest	Oldest	Oldest
MC/MC	Oldest (NN)	Oldest keeps trying but usually superseded by other NN	None

	No. of questions	No. of questioners	Question functions ( ) = downshiftings					Distribution of ques in discourse			Length of disc./mins.
			Open	Leading	Challenge	Sof. asser.	Clar.	Pres.	Expl.	Acc.	
NS/BC	10	Mainly 2	5	1	3(3)	1(1)	0	1	8	1	26
NS/MC	12	4	1	5	5	1	0	4	5	3	12
NN/MC	8	2	0	4(3)	3(2)	1	0	5	2	1	28
MG/MC	22	Mainly 2	3	1	7(7)	2	9	0	22	0	30

7. Implications for teaching

Given the variability of question functions and forms and the distribution of questions in discourse depending on the discourse structure and the interactional patterns, we agree with Candlin (in Coulthard 1977:xiii) that "learners need to become analysts of discourse themselves."

We would suggest that transcribed passages of task-oriented discourse from various English-speaking groups, including both native and non-native speakers in various combinations, along with the taped record, be studied by both ESL teachers and their students. We would envision as a first step the study of a passage which has been annotated in terms of discourse structure and selected language functions—for example, here this person is presenting reasons, here another is challenging someone, here they are exploring such and such an issue, here one person is trying to lead the others towards consensus, etc. A follow-up step would be to have students work closely with other, similar passages which have not been annotated, figuring out what interactions are going on between the participants, summarizing the ideas of various speakers and talking about why a particular participant asks, for example, a particular question and how that question and the response moves or does not move the substance of the discussion forward.

We would hope that in this way the language teacher and learner would gradually become sensitive to the range of possibilities that the language offers and the circumstances most commonly associated with the various options, and that in an informal and fairly indirect way, the learner would develop that systematic versatility which marks the communicatively competent person.

Oral and Written Syntactic Relationships in Second Language Learning

Roberta J. Vann

Iowa State University

This paper reports on a study which examined the oral and written language of second language learners. The study had two main objectives: to discover more about the relaionship between the oral and written syntactic units and structures of non-native speakers of English and to examine the reliability and utility of certain widely used indices of syntactic maturity. The experiment involved 28 Arabic-speaking adult subjects who were studying ESL. Comparable oral and written data samples were elicited by having subjects respond orally and in writing to a short film without narration. Transcribed data were analyzed using five indices of syntactic maturity. Teacher-evaluators also assessed oral and written data samples as a validation procedure. The indices were examined for their value as indicators of target language proficiency by comparing them with subjects' scores on a standardized test of English proficiency. Results suggest that second language learners tend to differentiate English oral and written expressive styles in ways similar to native speakers with written discourse containing fewer words, longer mean T-unit length, and fewer errors of certain kinds than comparable oral data. At the same time, this study casts doubt upon the usefulness of the T-unit and dependent clause ratio, indices which disregard learner errors, for use as measures of syntactic maturity of non-native speakers.

Although the link between oral and written language has been a topic of philosophical debate for centuries, this issue has elicited relatively few empirical investigations. The dearth is especially apparent in the area of applied linguistics. This is probably due, at least in part, to the until recently widely accepted notion of the primacy of speech and the consequent neglect of written language. For the pillars of linguistic thought the situation was clear: only speech was truly language and writing merely a means of recording it—certainly not a system in its own right. This attitude is exemplified by Wilhelm von Humboldt's statement which referred to writing as "only an incomplete mumified depository which needs, for full understanding, an imaginative oral reconstitution" (Salus, 1969:184). Other linguists who followed suit included Saussure (1959), Sapir (1921), Blomfield and Hockett (Vachek, 1973:10-11). Over the years, other linguists and psychologists have tended to agree with the Humboldt school, some to the extent of confusing speech with language and/or thought. This tendency drew the following comment from Vygotsky:

> We can trace the idea of identity of thought and speech from the speculation of psychological linguistics that thought is 'speech minus sound' to the theories of modern American psychologists and reflexologists who consider thought a reflex

> inhibited in its motor part. In all these theories the question of the relationship between thought and speech loses meaning. If they are one and the same thing, no relationship between them can arise. Those who identify thought with speech simply close the door on the problem (1962:2).

In a similar way, those who have reduced writing merely to speech written down also have attempted to shut a door on an important issue. In spite of notable exceptions to this conception of language (Vachek, 1973; Smith, 1975; Householder, 1971; and Bolinger, 1975) all of whom addressed the differences in oral and written language and dealt with writing as a system in its own right, it has been the view represented by behaviorists with their emphasis on verbal behavior and structuralists with their stress on speech as primary which has had a significant impact on foreign language teaching, particularly in the audiolingual approach in which the student is typically not allowed to write until after he/she has gained control over the structures orally. The assumption is that oral and written language exist linearly (oral preceeding written) with the implication that this is the natural developmental order for adult second language learners as well as for children acquiring their first language. Likewise, some versions of the direct method, in attempting to simulate the environment of first language acquisition, have maintained that speech in the target language be mastered before literacy skills are tackled. This rationale of natural order has been argued against by Ausubel who states that it is in fact ". . . unnatural to assume that after an individual becomes literate, he will learn in the same way as when he was illiterate" (1964:423). Ausubel's view is echoed by proponents of cognitive-code theory which encourages learners to use all their senses in assimilating material. Thus the cognitive-code approach, as did direct method, grammar translation, and audiolingual, makes an implicit assumption about the relationship of spoken to written language. Yet only a shallow empirical foundation for such assumptions exists.

Of the studies which have attempted to examine this relationship, most have been with children acquiring their native language. For that reason such work served as a basis for the present study. Inevitably research which examines syntactic relationships as does this one, must tackle the problem of finding an appropriate index of proficiency such as mean length of utterance (MLU), widely used by psychologists in studying first language learners, or the T-unit. The latter has been utilized by Hunt (1965) and others as a measure of syntactic maturity in English-speaking school children and later by second language learning researchers including Monroe (1975), Cooper (1976), Gaies (1977), Larsen-Freeman and Strom (1977) and others.

The controversy surrounding the matter of what constitutes an adequate measure and the problems involved in refining this index, although also the concern of the present project, are not the central focus of this paper. Rather, I am reporting here on the findings in this investigation which deal specifically with the relationship of the oral and written syntax of a group of adult second language learners at varying stages of their English language development. By collecting samples under carefully controlled conditions, the two products have been compared in an attempt to discover more about the relationship in itself as well as how it may relate to what we know about the oral/written development of L_1.

1. Subjects

Twenty-eight adult male Saudi Arabian subjects who had spent approximately five months in the United States and were enrolled in post-graduate courses in education, as well as ESL courses, provided the data used in this project. TOEFL scores, which were collected as a means of validating the language proficiency of the subjects thus enabling a realistic interpretation of the results of this study, as well as allowing for the possibility for future studies replicating this one, ranged from 337 to 570 with a mean of 420. The average subject was about 30 years old and had studied English in Saudi Arabia for about seven years. None of the subjects had ever previously lived in an English speaking country.

2. Procedures

In order to elicit comparable oral and written data, I sought an elicitation device which would naturally evoke a monologue. Largely for this reason, I chose a motion picture stimulus. This also follows the suggestion made by Scott and Tucker in an article concerning their study on oral/written relationships in [illegible] (1974:98) in which they stated that a film would be preferable to still pictures in encouraging subjects to be expressive. *Silences*, a 12 minute color film without narration has the added advantage of being of a subject matter, which while of adult appeal, is a film none of the subjects had seen. Nor had they been exposed to the plot—a story set during World War II in which a Yugoslavian peasant aids a wounded German enemy soldier. Because of its emotional intensity and thought-provoking final scene, the film evoked a strong interest in subjects and thus appeared to increase the likelihood of their wanting to respond to it. This was an important concern in choosing the stimulus. In the past, language researchers have sometimes neglected the need to present subjects with material which truly stimulates language production.

The subjects watched the film in groups of six. Following the film, three of the subjects were chosen at random to preceed immediately to one of three interview rooms where each subject told an interviewer who had not seen the film what the film was about and what his response to it was. Responses were recorded on an unconcealed tape recorder. At the end of approximately 10 minutes, interviewers sent the subjects to complete the writing phase of the experiment. Meanwhile, the other three subjects had remained behind and had been instructed to write a composition in which they told someone who had not seen the film what happened in it and their own opinion of it. These subjects were given approximately 20 minutes to complete the task before being sent to perform the oral phase of the experiment. By having half of the subjects write first while half spoke first, the possible effect of sequence could be examined.

Following data collection, oral recordings were transcribed orthographically, using conventional punctuation, then checked and verified by a phonetician who looked for possible discrepancies between the tapes and typed transcriptions. Finally, five minute segments were extracted and the transcriptions were re-typed. Written compositions were simply typed adhering to subjects' spelling, spacing and punctuation.

After excluding certain data considered extraneous, transcriptions were marked for T-units, mazes, dependent clauses and error-free T-units. The T-unit, the basic method of segmentation in this study, is defined by Hunt (1965:21) as each utterance grammatically capable of being considered a sentence. Thus all simple and complex sentences constitute one unit, while all compound sentences are sliced with the coordinate conjunction forming the first word of the second unit. Thus:

The man is a soldier/ and he wears a uniform (2 units)
The man is a soldier who is wearing a uniform (1 unit)

Words or strings which do not constitute T-units were labeled mazes. A maze, for the purposes of this study, is similar to what some researchers have referred to as a garble. Typical examples are semantic redundancies, unintelligible words or strings, and false starts. T-units were considered error-free if they made sense in the given context and were free from syntactic and lexical error. (In the case of written samples, spelling and punctuation were disregarded as criteria in determining whether or not a unit was error-free unless the spelling made a difference in meaning or grammatical correctness, e.g. *draw* for *drew* would have been tabulated as an error. This is closest to the procedure used by Scott and Tucker rather than Larsen-Freeman in similar studies. The implicit assumption here is that spelling and punctuation are "purely scribal conventions" (Hirsch, 1977:163) which probably occur at a relatively late stage in the writing process as the individual moves from an idea to a final graphological form.

3. Findings

In addition to computer analysis using the syntactic indices, the oral/written relationship was examined through holistic evaluations of the data by teacher-raters. Three highly-trained ESL teachers were asked to evaluate the oral and written data in 2 separate sessions according to their usual criteria for judging EFL composition. The procedure was a simple one: raters were asked to group data into 3 stacks according to the level of proficiency represented. As can be seen from Table 1, the raters tended to evaluate oral and written samples similarly to the way in which they had judged written ones. As can be seen from the high concentration of subjects along the left-right diagonal, those who were rated high, average or low one task were likely to retain this ranking for the other task as well.

TABLE 1

Raters' Groupings of Subjects on Oral and Written Assignments

Groups	I	II	III	Row Total
I	7	3	0	10
II	2	4	3	9
III	2	1	6	9
Column Total	11	8	9	28

Because judgments on the two sample groups were not truly independent in that the same raters judged each set, it cannot be concluded that good speakers make good writers or vice-versa. In fact, even if independence of rater judgments could be established, because the study was designed to disregard certain language features such as pronunciation in the oral samples and spelling and handwriting in the written ones, conclusions about transfer of proficiency from speaking to writing are necessarily limited.

The raters were also utilized in determining whether sequence of oral/written task influenced subjects' performances. Chi squared tests indicated that rater assignments on both oral and written assignments were independent of sequence. This finding shows that at least in terms of the conditions set up in this study, evaluators could detect no influence on a subject's oral performance as a result of his having written about it first, nor on his oral performance as a result of his having spoken about it previously for 10 minutes. This finding also allowed all subjects to be pooled together as one group in subsequent analyses.

These syntactic analyses revealed that the relationship between oral and written expression was apparent not only in the raters' holistic evaluations, but in terms of mazes, error-free T-unit length, and percentage of error-free T-units (see Table 2). This means, for example, that a subject's oral mean length of error-free T-units is likely to be reflected in his written mean error-free T-unit length. But the reader will note that for two of the indices, mean T-unit length and ratio of dependent clauses, there was no such correlation between oral and written performance. While it is possible that this result is due to the inad-

TABLE 2

Correlations between Oral and Written Indices

Indices	R	P
X T-Unit Length	.190	.16
% Dependent Clauses	.186	.17
% Mazes	.547	.00
X Length EF T-Units	.457	.00
% EF T-Unit/T-Units	.518	.00

equacy of the measures used with ESL subjects, it cannot be said with assurance that good speakers are good writers or vice-versa.

Since a portion of this study was devoted to examining these language proficiency indices, I decided to use multiple regression step analysis to determine which indices correlated with TOEFL scores, and whether the oral or written version was the better correlate in the case of each respective index. I also used multiple regression step analysis to determine the two best correlations with TOEFL scores, using first oral and then written data. I found that the oral ratio of mazes to words, the oral and written mean length of error-free T-units, and the oral ratio of error-free to total T-units correlate moderately with TOEFL scores. I also found that for use with the written sample, mean length of error-free T-units combined with ratio of error-free T-units is the most valuable discriminator, while for the oral sample, the ratio of error-free T-units combined with oral mean length of T-units accounts for the largest amount of variance. Note however that when combined with the previous index, it slightly lowered the amount of variance accounted for. Thus for written data, the error-free T-unit, especially its mean length, appears to be the best indicator of adult foreign language proficiency. It should be remembered that the 31 percent variance figure is deceptively low, since the TOEFL score represents an individual's overall English proficiency including listening comprehension, reading skills, and so forth. Thus even if a perfect measure of written syntactic proficiency existed, it would be unable to account for all the variance in the overall TOEFL score.

TABLE 3

Strongest Indices Correlating with TOEFL Scores for Written and Oral Data

Written				Adjusted
		Multiple R	R	R²
Step I	Mean length of Error-Free T-Units	.583 (p = .00)	.556	.310
Step II	Ratio of Error-Free T-Units to all T-Units	.607 (p = .00)	.555	.309
Oral				**Adjusted**
		Multiple R	R	R²
Step I	Ratio of Error-Free T-Units to all T-Units	.590 (p = .00)	.563	.318
Step II	Mean length of T-Units	.652 (p = .00)	.609	.371

In terms of the oral data, we note that ratio of error-free T-units to total T-units proved the strongest correlation with TOEFL scores, accounting for almost 32 percent of the variance when used alone. If oral mean length of T-units is added and used in combination with the ratio of error-free T-units, the amount of variance accounted for rises to 37 percent. Thus the mean length of T-units becomes a more useful index when combined with the ratio of error-free T-units. Again this suggests that syntactic length does function as a measure of language proficiency, but that errors need to be accounted for in order for it to show power as a discriminator.

In summary, those indices which take errors into account appear to be the best discriminators of language proficiency. It is interesting to note, however, that mean length of error-free T-units discriminates in terms of subject language proficiency similarly to the way in which mean length of T-units seems to discriminate according to the language proficiency of a native-speaker. Thus the length of a syntactic unit, when errors are taken into account, does have value as a discriminator of language proficiency for the individuals in this study, and results from other studies support the conclusion that this has wider application to other groups of language learners. (See Monroe, 1975 and Cooper, 1976).

As a follow-up study, the question was posed as to how raters had evaluated the oral and written samples in relation to each of the five oral and five written syntactic indices in this study. Would results be similar to those using the TOEFL as the criterion variable? In other words, would the same indices prove to be the strongest discriminators? This did, in fact prove to be the case, with mean length of T-units and percentage of dependent clauses failing to discriminate between the high and low proficiency groups at a statistically significant level. The other three indices, however, were statistically significant as discriminators.

This suggests that when raters are instructed to use their own criteria for evaluation, they tend to focus on a subject's errors in combination with his T-unit length in determining the level of his performance. On the other hand, the same raters do not appear to be substantially influenced by a subject's T-unit length or the percentage of his dependent clauses in his discourse. Once again, there is evidence that such indices, while functional in discriminating proficiency levels of native-speaker discourse, do not work effectively with non-native adults learning English, at least with the subjects in the present study.

In addition to the findings discussed thus far, certain aspects of the analysis resulted in information which lends itself to comparison with findings from first language acquisition studies—specifically those which have compared the oral and written discourse of English-speaking school children (O'Donnell, Griffin, and Norris, 1967; and Loban, 1976).

1) Oral compositions were almost twice as long as written ones, although oral segments represented only one-fourth the time allotment of written ones (5 minutes of oral taping versus 20 minutes of writing). The oral mean number of words produced was 422.9, while the mean number of written words produced was 214.8. These results are similar to those of other studies which have compared oral and written discourse. However, the discrepancy may be even greater with this group of ESL learners than with native-speakers.

2) Mazes were plentiful in speech and rare in writing. This is similar to Loban's finding with native-speaking school children.

3) Both mean length of T-units and mean length of error-free T-units were longer in written than in oral discourse. O'Donnell, Griffin, and Norris (1967:81) report that in this study with native-speaking children, mean length was greater in oral than in written expression until Grade 3 after which the trend reversed. Thus the subjects in the present study appear to have more in common with O'Donnell's post Grade 3 subjects.

4) The ratio of adverbial and adjectival clauses to T-units was higher in written than in oral language, especially in the case of adjectival constructions, the percentage of which more than doubled in written expression.

These findings suggest that the adult foreign language learners in this study, similar to native-speakers in previous studies, tend to use a more elaborative and more strictly controlled written than oral style. It is elaborative in the sense that T-units are longer with a larger percentage of adverbial and adjectival clauses and more strictly controlled in that fewer words are produced in a given time allotment with mazes representing a very small proportion of these words. Still open to speculation, however, is the question of how much this is a function of language training—in the case of the subjects in this study training in English in a Western country which persumably exposed students to a more elaborate and finely controlled written style as opposed to a more free oral style characterized by shorter syntactic units.

Yet, for whatever reasons, it does appear that evidence from the oral/written language relationship of these subjects offers some support for the similarity of first language acquisition and second language learning even when the two groups are widely removed in age, native language, and cultural background.

4. Implications

While lending additional support to the notion that second language learning closely approximates the developmental sequence of first language acquisition, the present study also opens up a vast range of further research questions from the merely interesting (e.g. how universal might this phenomenon be?) to the pragmatic and pedogogical (e.g. how might we help good speakers become better writers?).

Unfortunately, we are not likely to discover the answers to such questions until we have developed better methods of data evaluation—both through holistic rating and objective means. Thus the goal of developing more refined methods of measuring language proficiency must be among our highest priorities in future language research.

Short-Term Changes in EFL Composition Skills[*]

Bradford Arthur
University of Michigan

This paper describes the results of a computer-assisted analysis of 152 compositions written over a period of seven weeks by 14 lower-intermediate level EFL students. Measurements on these compositions estimate changes in grammatical complexity, grammatical error frequency, grammatical error type, fluency, vocabulary size, spelling, and punctuation. Such measurements provide an objective look at some short-term changes in the writing skills of adult learners of English.

Single compositions are unreliable representations of writing proficiency, but comparisons of average measurements on several compositions reveal a number of changes in composition skills. Several such changes depend on the first language and culture of the learner. A comparison of composition changes with changes in placement test scores indicates that some important developments in writing are not reflected by standard discrete-point tests. A comparison of measurements for eighteen compositions with ESL teachers' subjective ranking of those same compositions shows that teachers consistently base their rankings on the same limited set of criteria.

Over the past several years I have taught writing skills to adult foreign learners of English. Like most such teachers, I have been aware of substantial differences of ability among the students in each class. Some were atrocious spellers; others spelled English accurately. Some made many grammatical errors, while others were able to eliminate or avoid virtually all such errors. Some organized their writing in a way that matched my American sense of order; others wrote in ways that seemed to me disjointed and illogical. Again, like most teachers of writing, I had an intuitive sense that I was helping my students to improve their writing. Although all students had off-days, I sensed some improvement between the beginning and end of an eight-week intensive-course session. But until last year, I had no concrete evidence to support my intuitions. Changes that I sensed were subtle rather than dramatic. The bad spellers were still bad spellers, but maybe not quite so bad. The same grammatical errors occurred, but perhaps not quite so often.

About a year ago, when I was teaching a class of 14 lower-intermediate adult students of English as a foreign language, I decided to measure, with as much accuracy and objectivity as possible, some of the changes that occurred in the writing of the students in my class over a period of seven or eight weeks. What follows is a description of those changes. This description is based on a computer-assisted analysis of 152 individual compositions. For each of these compositions I was able to obtain an approximation of vocabulary size, frequency of errors in grammar, spelling, and punctuation, writing speed, and the grammatical complexity of sentences. One of my objectives was to determine whether any of these measurements suggested changes in student performance over the eight weeks of the course. Another objective was to determine whether these measurements were related to the sorts of intuitive judgments that composition teachers make about their students' writing and also whether changes in these measurements corresponded with changes in standard language-achievement test scores.

---

[*] I am grateful to Ann Borkin for suggestions that led to several revisions.

1. Method

The students. The information presented here is, I believe, an accurate representation of some changes in my students' written English, but you must decide whether for your purposes my students should be considered "typical". In age, almost all were in their late teens or twenties. Only one student in the class was in her thirties. The class represented a variety of different countries in Latin America and the Middle East; one student was from Japan. All but three of the students spoke either Spanish or Arabic. They had come to the United States to study English in preparation for study at an American university.

Students were placed in my class based on an objective test measuring their grammatical knowledge, their knowledge of English vocabulary, their reading and oral comprehension. This test did not measure writing ability directly. When my students were given a similar test at the end of the eight-week session, all but one had improved enough to be advanced to the next course level.

During the term, students wrote twelve timed in-class compositions. Topics for these compositions alternated between descriptions of short animated films and other topics based on the discourse strategy treated during that week—chronological order, cause-and-effect, and so on. These compositions, a total of 152, constitute the data for my analysis.

The measurements. Table 1 describes the nine measurements made on each of the 152 compositions.

The average number of words per minute of writing time is assumed to reflect the writer's fluency. Because I saw the care that each student gave to his or her writing, I am confident that writing speed does not simply measure the students' ability to spew out minimally coherent sequences of English words. The students were never pushed to write more or to write faster. They took pains to write carefully.

The length of T-units—that is, of single independent clauses together with all modifying subordinate clauses—was proposed by Hunt (1965) as a measure of syntactic complexity. T-units are generally lengthened by sentence embedding or by the ellipsis of repeated elements in compound sentences. Hunt found that T-unit length reflected the verbal maturity of school-age writers.

TABLE 1

The Nature of the Measurements and the Abilities Measured

measurement	*ability of writer*
1. Average words per minute of writing time	fluency, articulateness
2. Average length of T-units	grammatical sophistication
3. Percentage of error-free T-units	grammatical sophistication
4. Average length of error-free T-units	grammatical sophistication
5. Type-token ratio	size of active vocabulary
6. Average number of grammatical errors per 100 words	ability to correct or avoid grammatical errors
7. Average number of misspelled words per 100 words	ability to spell English correctly
8. Percentage of T-units ending with a punctuation mark	ability to punctuate written English appropriately
9. Semantic errors as a percentage of total errors	seriousness of grammatical errors

Since adult learners of English often write long T-units by translating sentences from their native language word-for-word into English or by using interlanguage rules to combine English words, longer T-units in the writing of foreign adults may not indicate a more sophisticated control of the grammatical constructions of English. To overcome this difficulty Scott and Tucker (1974) use a different measure of grammatical sophistication: the percentage of T-units that are free of grammatical errors.

Larsen-Freeman (1978) describes yet another measure of the grammatical sophistication of foreign learners of English. Like Hunt, she calculates the average length of T-units, but unlike Hunt, she includes in her calculation only those T-units that are error free.

A type-token ratio is frequently used to estimate active vocabulary size. The variant of that measurement used for this study is the number of word types (i.e. different words) divided by the square root of twice the total number of words in the text. Thus, if the writer used the same limited number of words over and over again, the type-token ratio for his/her composition would be a relatively small number; whereas the type-token ratio for a writer using a wider variety of words would be a larger number.

The frequency of grammatical errors in each composition was measured as the average number of errors per 100 words of text. As in any error analysis, these are errors from the viewpoint of a native speaker. Presumably if the learner-writer had perceived them as errors, he or she would have corrected them.

Spelling proficiency was estimated by the number of misspelled words per 100 words of text.

A rough measure of the writer's capacity to punctuate English sentences was provided by calculating the percentage of T-units ending with a punctuation mark. (Virtually all T-units require some sort of terminal punctuation.) This measurement, however, does not reflect the writer's ability to provide necessary punctuation within T-units, nor does it take into consideration the appropriateness of the particular punctuation mark used at the end of each T-unit.

Some errors distort or obscure the meaning of a sentence. Such an error is committed, for example, when a learner writes the verb *killed* with the intended meaning *died* or uses present tense in reference to past events. On the other hand, when a writer fails to make the subject and verb agree or puts an adverb in the wrong place, or hopes *to going* rather than hoping *to go*, he proclaims his foreignness, but does not distort his meaning. Semantic errors, errors that distort meaning, are more serious. Consequently a decrease in the ratio of semantic errors to total errors indicates an improvement in the quality of the learner's errors.

Each measurement in Table 1 is associated with an ability generally considered important in the writing of English. Note, however, that these abilities are not measured directly. Such things as fluency and grammatical sophistication can not be counted or calculated. Consequently, references below to changes in these abilities are only as valid as the association between each ability and the corresponding measurement or measurements. Note in addition that some of the most important concerns of writing teachers are not included—such things as organization, relevance, tone, and cohesion. These are missing because I was unable to measure them objectively, and such omissions limit the scope of this analysis of student writing.

2. Discussion

The procedure that revealed changes in composition skills most clearly turned out to be one of the simplest. The twelve compositions were divided into two groups: those written during the first half of the course and those written during the second half. For each half, the individual composition scores for each student were averaged. Then these average scores for each of the fourteen students in the class were themselves averaged, arriving at a single set of average or mean scores for the first six compositions and another set of average scores for the last six. These scores are presented in Table 2.

There was clearly significant improvement in three of the nine measurements comparing the first six compositions with the last. A significantly higher percentage of T-units were punctuated in later compositions. As indicated by the type-token ratio, students used a significantly larger vocabulary in writing the later compositions. Finally, later compositions were written slightly more rapidly, an increase in writing speed averaging about one word per minute.

Two other measures were within five percent of the accepted .05 significance level, and therefore suggest some other likely changes in the writing of my students. The level of misspelling decreased by roughly .6 words per 100, and between three and four percent more of the T-units in the later compositions were error free.

There was no significant change in the frequency of grammatical errors.

TABLE 2
First Half—Last Half Comparison of Average Values for Nine Measurements—All Students

measurement	first 6 compositions	last 6 compositions	difference (last—first)	significance of difference
1. errors per 100 words	17.35	16.77	−.58	
2. average words per T-unit	10.56	9.99	−.57	
3. error-free T-unit frequency	25.21	28.89	3.68	c
4. average words per error-free T-unit	6.43	6.23	−.20	
5. terminal punctuation frequency	80.89	85.02	4.13	b
6. type-token ratio	4.31	4.47	.16	a
7. writing speed	5.84	7.02	1.18	a
8. frequency of misspelling	7.66	7.04	−.62	c
9. semantic error frequency	36.24	34.63	−1.61	

(N = 14)

One-tailed paired T-test
[a] $p < .01$
[b] $p < .05$
[c] $p < .1$

This result surprised me since a major portion of class time was spent in error-related exercises and since students were required to rewrite their compositions correcting grammatical errors. Again contrary to my expectations, the average length of T-units did not increase. If anything, it seems to have decreased since the average length of all T-units and the average length of error-free T-units both decreased slightly.

Among the students in my class there were six Latin American students whose native language was Spanish and five Near-Eastern students who spoke Arabic as a native language. It was therefore possible to compare the performance of these two groups over the course of an eight-week period when they were receiving virtually identical training in English.[1]

Table 3 compares the average proficiency level of these two groups of students on each of the nine tests. The results are not surprising. Spanish speakers were significantly better at spelling and punctuating English—almost certainly due to the consistent similarities between the writing systems of Spanish and English and the equally consistent differences between the graphic conventions of Arabic and English. Spanish speakers also used a larger vocabulary in their writing, as one would expect given the vast number of English

---

[1] This training included four intensive courses (reading, writing, speaking and listening), a total of 20 hours per week. This particular group of 14 students worked together in all four courses.

TABLE 3
Comparison of average Scores for Arabic and Spanish speakers on nine measurements

measurement	Native language: Arabic	Spanish	Significance of difference
1. errors per 100 words	16.28	17.41	
2. average words per T-unit	10.63	10.49	
3. error-free T-unit frequency	26.08	27.68	
4. average words per error-free T-unit	6.21	6.25	
5. terminal punctuation frequency	72.86	86.13	b
6. type-token ratio	4.18	4.56	b
7. writing speed	6.12	6.86	
8. frequency of misspelling	13.18	3.84	
9. semantic error frequency	33.24	36.45	b
N:	(5)	(6)	

Two-tailed T-test
$^{b}p < .05$
Variances for the measurement of misspelling frequency are grossly unequal. Two Arabic-speaking students had severe spelling difficulties; all other students were reasonably good spellers.

cognate and loan words in Spanish. On all of the other tests, the performance of Arabic and Spanish speaking students was not significantly different.

But the changes in writing skills that occurred during the eight-week course were strikingly different. These differences are represented in Tables 4 and 5. As Table 4 indicates, significant changes in the writing skills of the Spanish-speaking students were their increased writing speed and their increased capacity to avoid semantic errors. Also notable were improvements in spelling and in the frequency of error-free T-units. However, there was a possible decrease in the average length of error-free T-units. It is possible that Spanish speakers in the class were able to write more error-free T-units partly because they were writing shorter T-units.

As Table 5 indicates, none of the changes in scores of Arabic-speaking students were similar to changes among the Spanish-speaking students. Instead, Arabic-speaking students improved significantly in their ability to punctuate English sentences and in the size of their active vocabulary. The writing of both groups changed, but in quite different ways.

Characteristics of the learner's first language have long been considered a source of his or her generalizations about a second language and consequently a source of learner errors. The present data suggest that differences in language or culture (I cannot determine which) also affect the direction and nature of the language learning process. Different cultural groups may require different

TABLE 4

First half—last half comparisons of average values for nine measurements—Spanish students

measurement	first 6 compositions	last 6 compositions	difference (last—first)	significance of difference
1. errors per 100 words	17.91	16.93	−.98	
2. average words per T-unit	10.87	10.07	−.80	
3. error-free T-unit frequency	25.44	30.17	4.73	c
4. average words per error-free T-unit	6.58	5.91	−.66	
5. terminal punctuation frequency	84.29	88.30	4.01	
6. type-token ratio	4.49	4.64	.15	
7. writing speed	6.21	7.54	1.33	b
8. frequency of misspelling	4.13	3.54	−.59	c
9. semantic error frequency	39.43	33.39	−6.04	a
(N = 6)				

One-tailed paired T-test

[a] $p < .01$

[b] $p < .05$

[c] $p < .1$

TABLE 5

First half—last half comparisons of average values for nine measurements—Arabic students

measurement	first 6 compositions	last 6 compositions	difference (last—first)	significance of difference
1. errors per 100 words	16.16	16.35	.19	
2. average words per T-unit	11.02	10.26	−.76	
3. error-free T-unit frequency	23.83	28.50	4.67	
4. average words per error-free T-unit	6.28	6.18	−.10	
5. terminal punctuation frequency	68.50	76.76	8.26	b
6. type-token ratio	4.11	4.26	.15	b
7. writing speed	5.73	6.62	.89	
8. frequency of misspelling	13.86	12.54	−1.32	
9. semantic error frequency	31.97	33.65	1.68	
(N = 5)				

One-tailed paired T-test

[b] $p < .05$

curricula to compensate for differing tendencies to develop in some writing skills but not in others.

To determine the consistency of teacher judgments in the ranking of compositions, I compiled a set of eighteen compositions all written on the same subject by students in my class. These compositions were given to a total of 13 ESL teachers and teachers-in-training who were asked to rank the compositions from best to worst. No specific criteria for making judgments were suggested; tie rankings were accepted.

These teachers were surprisingly consistent in their rankings; there was general agreement as to the ordering of compositions roughly 60 percent of the time.[2] The teachers themselves were surprised by their consistency; they found very difficult the task of ranking compositions so similar in quality. Yet all of the teachers seemed to be using roughly the same criteria in judging compositions. What were those criteria? More particularly, were any of the writing skills considered important by teachers captured by the the nine measurements used for this study?

In fact, three of the nine measurements were related to the rankings given by the teachers; they are, in decreasing order of importance: writing speed, frequency of grammatical errors, and frequency of spelling errors.[3] There was a correlation of .91 between these three measurements considered together and the average teacher-rankings. By far the single most important predictor of teacher rankings was writing speed, that is, the length of each composition. It would be uncharitable, however, to assume that composition teachers were judging quality simply by length. I suspect that students who could write faster also wrote more articulately, and that it was this less tangible characteristic to which teachers were attending.

Because all of the compositions ranked by the teachers were written on the same topic, it was possible to obtain at least a rough measurement of the relative amount of information conveyed by each writer. I guessed that this measurement, in addition to the original nine, might be associated with the judgments made by teachers in ranking compositions, but in fact it was not. Apparently teachers, in making their judgments, were attending more to the form of writing than to the completeness of its content.

In ranking papers written by my students, teachers were attending to characteristics associated with length and with the frequency of grammatical and spelling errors. It seems that the sorts of changes that occurred in the writing of the Latin American students were valued more highly by writing teachers than the sorts of changes that occurred in the writing of Near-Eastern students. One might expect that these teachers would consider Latin American students better writers of English than Arabic students. But such generalizations are dubious. Had the compositions themselves been different—written on a different topic or by students more divergent in ability—the criteria used by teachers to evaluate them might also have been different.

---

[2] Based on an analysis of principal components using individual teacher rankings as variables.

[3] Based on a statistical procedure that selects independent variables for a regression model.

Although teachers were surprisingly consistent in their judgment of my students' compositions, my students were inconsistent in the quality of their writing. As a consequence, a single composition did not provide reliable evidence of a student's writing ability. Figures 1, 2 and 3 illustrate the problem. These three graphs show the range of scores over twelve compositions for a single student on each of the three measurements most closely associated with teacher judgments, that is, speed, grammatical error frequency, and frequency of misspelling.

Although a comparison of the writing speed of the first composition represented in Figure 1 with the speed of the last would suggest a dramatic change in writing speed, this comparison would be misleading. Looking at the pattern of change over all of the compositions, it is difficult to decide whether speed has increased. Ignoring the first and last compositions, it would almost seem that speed is decreasing. Similarly, the drastic drop in error frequency in Figure 2 between the first and last composition does not, as the rest of the graph shows, indicate a gradual decline in the number of errors made by this writer. The decrease in errors is only apparent in the last two compositions. In Figure 3, the fluctuation among compositions totally obscures any gradual change that might be occurring in spelling proficiency.

This sort of evidence casts doubt on the practice of evaluating a student's writing proficiency based on a single composition or of evaluating changes in student writing by considering a single pair of compositions. Unfortunately, accurate evaluation of student writing skills seems to require a larger set of compositions as evidence.

Students were placed in my class on the basis of a multiple-choice language proficiency test measuring grammatical knowledge, vocabulary size, and comprehension of written and spoken texts. It is often assumed that such tests, although they do not measure writing skills directly, do indirectly indicate the level of writing skill. My analysis of student compositions gave me a basis for testing that assumption. In particular, I was able to determine which if any of the changes in writing scores proved to be good predictors of changes in proficiency test scores. In fact, I discovered that of all the nine measurements only vocabulary size was significantly related to scores on the proficiency test.[4] Since part of the proficiency test specifically measured knowledge of English vocabulary, this association is not surprising. What is surprising is that none of the three measurements most valued by composition teachers was associated with proficiency test scores. Students with very similar proficiency test scores may differ in their mastery of important writing skills, and consequently when proficiency tests are used to place students in ESL classes, writing teachers are likely to have heterogeneous classes.

---

[4] Based on a statistical procedure that selects independent variables for a regression model.

FIGURE 1

Changes in an individual student's writing speed (in words per minute) over 12 compositions

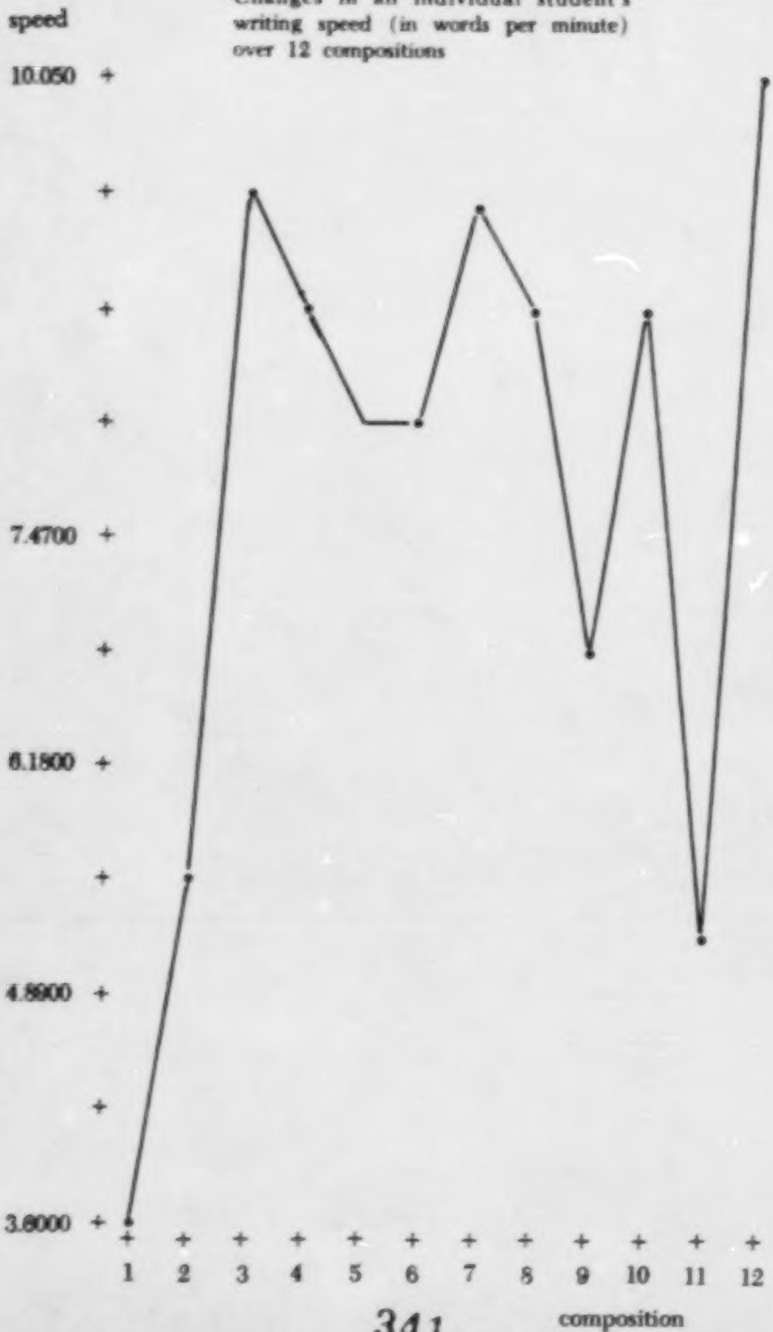


FIGURE 2
Changes in an individual student's grammatical error frequency (errors per 100 words) over 12 compositions

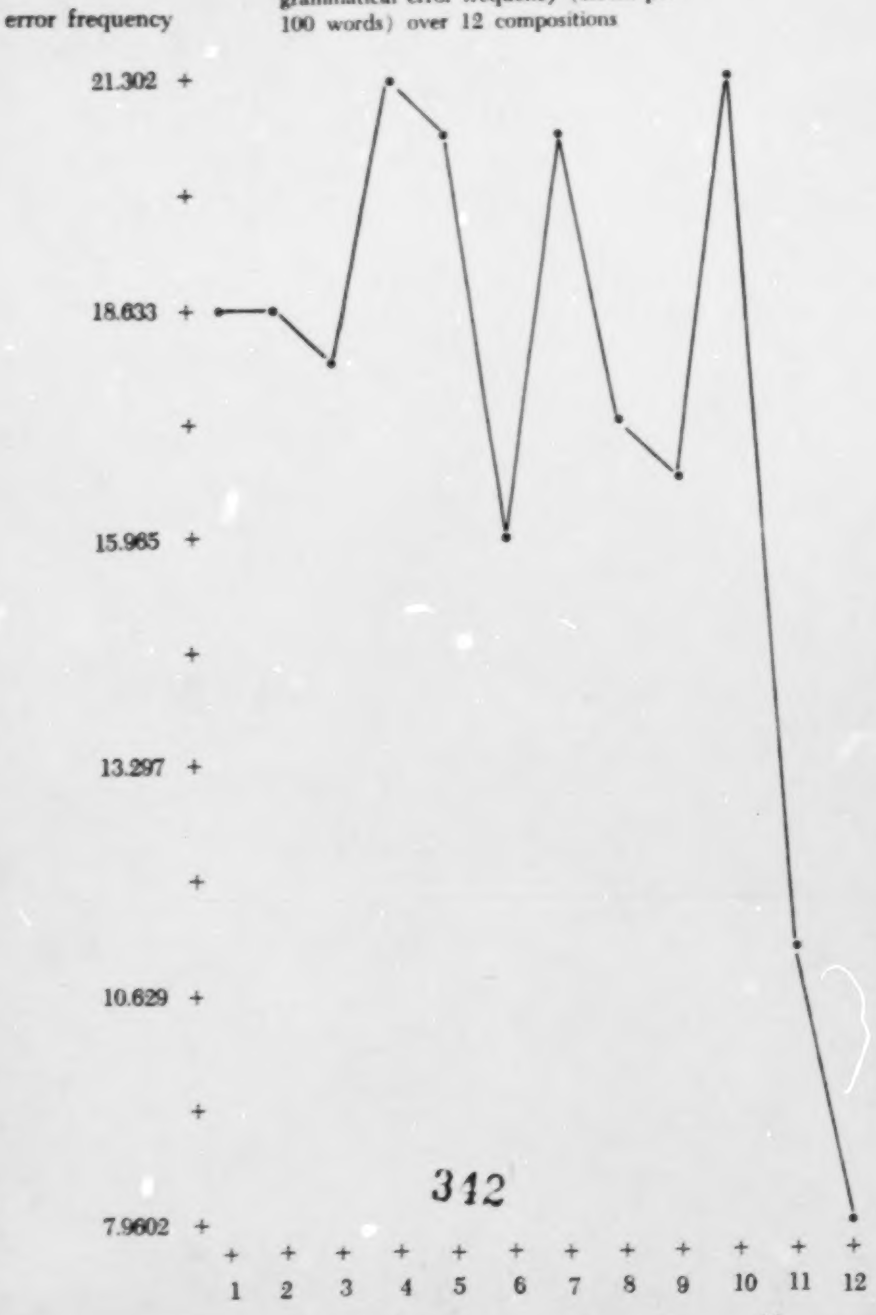
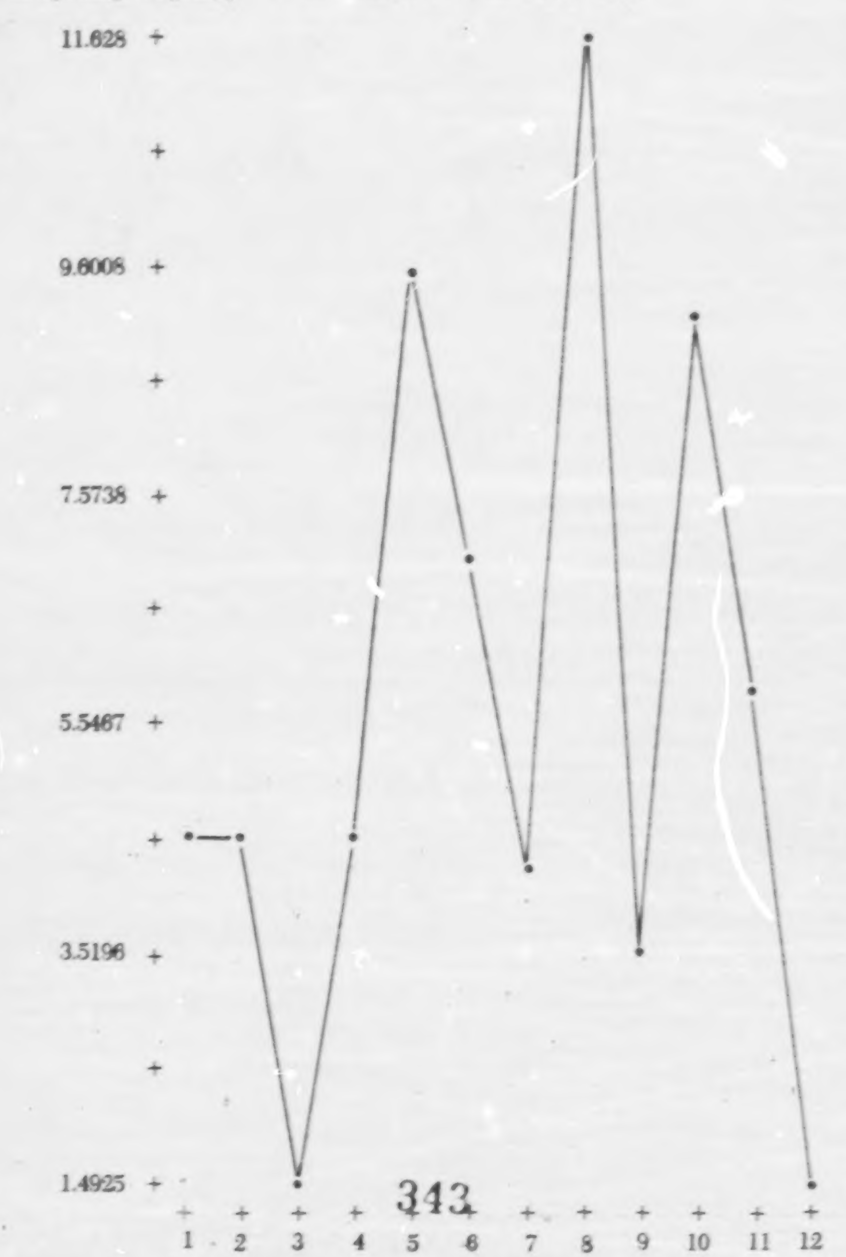


FIGURE 3

Changes in an individual student's misspelling frequency (misspellings per 100 words) over 12 compositions



In summary, the study of my students' compositions led me to five conclusions. First, my students did improve with regard to some of the writing skills I was able to measure. The most notable improvements were in writing speed and in vocabulary size. Second, the writing of students in my class from different language and cultural backgrounds changed in different ways. These differences were due, I believe, to differences in the knowledge, experience, and expectations that the students brought to the class rather than to differential treatment within the class itself. Third, ESL teachers were surprisingly consistent in their evaluation of compositions written by my students. Moreover, they based their evaluations primarily on criteria related to writing speed, frequency of grammatical errors, and frequency of misspelling. Fourth, a single composition provided an unreliable basis for judging the writing ability of an individual student. Similarly, comparing two individual compositions was an unreliable procedure for evaluating change in composition skills over a period of seven or eight weeks. Finally, the objective tests used to evaluate the language proficiency of my students and to place them in language classes did not measure their proficiency as writers.

These conclusions are based on a single group of students all within a relatively narrow range of English proficiency, all roughly the same age and all studying together in the same ESL intensive course. They are in no sense a random sample of second language learners of English. Groups with different characteristics might not behave in the same way. I believe, however, that my study and my conclusions may be useful in directing further inquiry, and I suspect that at least some of my conclusions may prove to be valid in classrooms other than my own. I hope, moreover, that my study will encourage others to continue the development of techniques for measuring short-term changes in the performance of language learners. Only when such measures exist can teachers rationally evaluate the success of their teaching.

Syntactic Skill and ESL Writing Quality

Patrick T. Kameen
Syracuse University

A more thorough understanding of the relationship between syntactic skill and ESL writing quality—an understanding based on a solid body of empirical data rather than on vague intuitions—will better prepare composition teachers to help their ESL students learn to write.

In this paper suggestions are made for replacing vague intuitions with objective measures of syntactic skill, with these suggestions based on the results of a recent exploratory study analyzing 40 syntactic factors in 50 compositions (25 "good" and 25 "poor" compositions) written by college-level ESL students.

The differences between the two groups of writers were simultaneously statistically significant at the .05 level in terms of (1) T-unit length, (3) clause length, and (3) incidence of dynamic passive constructions.

The implications of these findings are discussed and suggestions are made for the use of such statistical counts of syntactic factors, the wider use of sentence combining exercises, and a re-examination of present attitudes toward the passive voice.

Intuition, rather than empirical evidence, has long been the basis upon which composition teachers have judged the writing skill of their ESL students. Unfortunately, the intuitions handed down through the years from rhetorician to composition teacher offer few insights into the relationship between writing quality and syntactic skill (see Hunt 1977).

The purpose of this exploratory study was to determine if there is a correlation between syntactic skill and scores assigned to compositions written by college-level ESL students.

1. Method

Subjects. A randomly-drawn sample of 50 compositions written by college-level ESL students (25 "good" and 25 "poor" writers) at Syracuse University was used. Fourteen different native languages were represented in the sample (see Table 1).

Task. Each composition analyzed was a 30-minute timed writing on an assigned topic given as part of the Michigan Test of English Language Proficiency. Because the test was administered a number of times in 1976, several different topics were used (see Table 2).

Procedure. Each composition had been graded holistically on a scale of 100 by two highly-qualified, independent raters with over twenty-five years of experience in grading compositions written by ESL sudents. This researcher did not grade any of the compositions. In no case did the raters disagree by

TABLE 1
Native languages of subjects

Language	"Good" Writers	"Poor" Writers
(1) Arabic	10	10
(2) Chinese	1	0
(3) Farsi	1	4
(4) French	1	0
(5) Hebrew	1	0
(6) Ibo	2	0
(7) Italian	1	0
(8) Japanese	0	1
(9) Lambya (Malawi)	2	0
(10) Malay	3	0
(11) Portuguese	0	1
(12) Spanish	2	7
(13) Thai	0	2
(14) Urdu	1	0

Task—Each composition analyzed was a 30-minute timed writing on an assigned topic given as part of the Michigan Test of English Language Proficiency. Because the test was administered a number of times in 1976, several different topics were used (see Table 2).

more than 3 points which strongly suggests that they were applying similar standards in evaluating the compositions. Small differences in grades were averaged to arrive at a single, final grade based on the raters' general impressions of its quality. Good compositions were those that had been assigned grades of 79 and above, with a mean of 80.6. Poor compositions had received scores of 68-71, with a mean of 70.2.

The compositions were all of different lengths and were read in the form that the students had submitted them. Physical appearance, then, may have contributed to a halo effect—the tendency to assign a score partially on the basis of the writer's more appropriate use of the conventional mechanics of writing, such as punctuation and spelling (see Braddock, Lloyd-Jones, and Schoer 1963).

Non-parametric (distribution-free) procedures were used to statistically analyze the data.[1] The Wilcoxon test, one of the most powerful rank tests, was used to convert raw scores to z scores—their respective number of standard deviation units from the mean. In general, this study analyzed such things as length of various types of writing units (e.g. T-units[2], clauses, and sentences), incidence of passive voice,[3] types of clauses, and types of joining devices be-

---

[1] I am indebted to Arthur Roth and Kenneth Kaminsky, Syracuse University statisticians, for their help with the statistical analysis.

[2] A T-unit is "one main clause plus all the subordinate clauses attached to or embedded within it" (Hunt 1965:49). Thus, *The men are standing on the corner, and they are laughing* is considered two T-units, with the first T-unit containing 7 words, ending at the comma, and the second, containing 4 words, beginning with and including the conjunction *and*. However, in *The men who are standing on the corner are laughing*, there is only one T-unit consisting of 10 words.

[3] Passive voice is defined here as the true or dynamic passive as opposed to the stative passive. Examples, such as the following, were counted: *They were given the notice* and *He had been told to sit down*. The presence or absence of a *by*-phrase had no bearing on those passive constructions chosen for use in this study. Stative passive examples, such as the following, were not counted: *I am interested in the results* and *My coat is torn*. Such distinctions were made in other studies to which this study is compared.

TABLE 2

Topics Assigned

Topic	"Good" Writers Writing on Topic	"Poor" Writers Writing on Topic
(1) The U.N. should pass a law limiting a woman to a maximum of 3 children or 3 pregnancies (whichever comes first). Do you agree or disagree? Explain your answer.	15	14
(2) Who do you consider the most important person who has ever lived? Give a short biography of the person and explain why you have chosen him/her.	2	4
(3) Which event in history has had an important effect on your country? Describe this event and how it has changed and/or influenced the progress and development of your country.	4	3
(4) All men and women should be required to serve their country either in the army or in some other government-designated way for a minimum of 2 years. Do you agree or disagree? Present arguments to support your position.	1	1
(5) What is the value of education to the individual and to society in your country? Give examples, reasons, and supporting arguments in your composition.	2	1
(6) Do you believe that honesty is a quality that government officials should have? If so, why? If not, why not?	0	1
(7) International students who are sponsored have an obligation to return to their own countries after they have finished their training in the U.S. Do you agree or disagree? Develop your views giving reasons and supporting arguments.	1	1

tween writing units.

A two-sided rejection region of one half the test statistic was used, and the criterion figure for significance within this level of confidence was ±1.96. The Bonferroni inequality test, with a criterion figure of ±3.025, was then applied to determine if good writers in the sample differed from poor writers in terms of simultaneous factors—factors which form a cluster and which, when taken together, characterize the writing of a certain group.

2. Conclusions

The Wilcoxon test revealed that 15 of the 40 factors were significant as individual factors differentiating between the writing of good and poor writers.

The Bonferroni inequality test revealed that 12 of these factors were simultaneously significant. Twenty-five factors were nonsignificant, and the 12 simultaneously significant factors naturally collapsed into three larger, more general categories: (1) T-unit length, (2) clause length, and (3) incidence of passive voice. Table 3 displays the statistical results for the 40 factors considered in this study.

TABLE 3

Syntactic factors

Factor	"Good" Writers Mean	"Poor" Writers Mean	z score for "Poor" Writers
A. Simultaneously significant factors:			
1. Number of long T-units (21+ words)	5.60	2.24	−4.71
2. Number of dynamic passives	3.40	.50	−4.69
3. Number of sentences with at least 1 passive	3.10	.50	−4.65
4. Number of words in non-fragment sentences	294.20	203.40	−4.49
5. Passives per clause	.12	.03	−4.47
6. Words per clause	10.83	8.54	−4.45
7. Passives per T-unit	.20	.05	−4.39
8. Passives per non-fragment sentences	.24	.06	−4.38
9. Words per T-unit	18.40	14.30	−3.62
10. Number of passives in main clauses	2.16	.24	−3.50
11. Ratio of long T-units to all T-units	.36	.19	−3.41
12. Number of passives in secondary clauses	1.24	.28	−3.12
B. Individually significant factors:			
13. Ratio of short T-units to all T-units	.11	.26	2.77
14. Number of words per non-fragment sentences	23.00	18.30	−2.50
15. Number of short T-units (1-8 words)	1.90	5.04	2.16
C. Non-significant factors:			
16. Number of secondary clauses	11.00	9.10	−1.48
17. Number of words in fragments	1.30	4.20	1.39
18. Number of clauses	27.30	24.80	−1.23
19. Number of adverbial clauses	4.44	3.64	−1.21
20. Number of T-units joined by *and, but, or*	2.04	2.76	1.14
21. Number of non-fragment sentences	13.40	12.10	−1.13
22. Number of noun clauses	3.50	2.76	−1.01
23. Ratio of #20 to all T-units	.12	.16	0.87
24. Number of actives in secondary clauses	9.76	8.70	−0.82
25. Ratio of #26 to all sentences	.03	.01	−0.78
26. Number of sentences begun with *and, but, or*	.40	.20	−0.72
27. Number of T-units (= main clauses)	16.30	15.70	−0.72
28. Clauses per T-unit	1.74	1.66	−0.67
29. Ratio of secondary to all clauses	.40	.37	−0.62
30. Number of actives in main clauses	14.10	15.30	0.62
31. Ratio of adjectival to all secondary clauses	.28	.33	0.59
32. Number of fragment sentences	.04	.20	0.48
33. Ratio of nominal to all secondary clauses	.29	.27	−0.48
34. Ratio of mid-length to all T-units	.53	.55	0.47
35. Number of adjectival clauses	3.04	2.68	−0.29
36. Number of mid-length T-units (9-20 words)	8.84	8.44	0.16
37. Number of T-units joined by (;)	.72	.60	0.14
38. Ratio of adverbial to all secondary clauses	.42	.40	−0.12
39. Ratio of #37 to all T-units	.04	.04	0.09
40. Number of actives	23.90	23.90	0.03

Clearly, there is a correlation between incidence of certain syntactic factors and scores assigned to compositions written by college-level ESL students, thus supporting the alternative hypothesis and rejecting the null hypothesis. The most interesting findings involve T-unit length, clause length, sentence length, and incidence of passive voice.

T-unit length, especially incidence of long T-units (21+ words), is apparently a powerful index for differentiating between the writing of good and poor college-level ESL writers. Good writers wrote approximately 29% more words in each T-unit, writing an average of 18.40 words/T-unit as opposed to 14.30 for poor writers.

This finding correlates with the findings of three other researchers. Hunt (1965), in his analysis of the writing of 4th-, 8th-, 12th-graders and professional adults, found T-unit length to be the most reliable index of syntactic maturity through the grades. Potter (1967), who analyzed the writing of good and poor 10th-grade writers, found that good writers wrote longer T-units than poor writers. In addition, Schmeling (1969), in his analysis of college freshman writing, found that there was a significant difference in T-unit length between first and improved versions of quality-differentiated compositions.[4]

Mean clause length, found by Hunt (1965) to be the second most powerful indicator of syntactic maturity, correlated significantly with rated quality of writing among the college-level ESL writers in this study, but not among the college freshman writers in Schmeling's (1969) study. Of equal importance, however, the results of this study indicate that there were no significant differences between good and poor college-level ESL writers either in terms of number of clauses attached to or embedded within a T-unit (clauses/T-unit) or in terms of mean incidence of various types of clauses—nominal, adjectival, or adverbial. These findings have widespread implications, pointing us in the direction of the types of exercises from which "poor" writers may benefit, and away from those from which "poor" writers will most likely not benefit.

First, the commonly-held intuition that "good" writers have a superior command of the use of subordinate clauses, allowing them to embed more clauses of various types within a main clause matrix than do "poor" writers, is in no way supported by this study, thus inviting us to question the practice of emphasizing, in the ESL composition classroom, the mastery of the techniques of subordination. Obviously, "good" writers earned their higher quality ratings and "poor" writers their lower quality ratings for reasons other than their ability to write subordinate clauses.

Second, since it was clearly an increase in the *number of words* in each clause, rather than an increase in the *number of clauses*, that accounted for the fact that "good" writers wrote significantly longer T-units than did "poor" writers, "poor" writers may benefit from exercises in sentence combining. And with sentence combining exercises—exercises which present the students with numerous short sentences directing them to combine them into longer, more economical units (see Kameen 1978)—we can give students practice in writing longer clauses and T-units, particularly by showing them how to reduce full *clauses* to prepositional, infinitival, and participial *phrases*. By our showing "poor" writers the various techniques for reducing and consolidating larger and larger chunks of information into fewer (but longer) clauses, we will be leading these students in the direction of reduced redundancy and increased succinctness (see Hunt 1965), thus helping "poor" writers write more like their "good" writing counterparts.

---

[4] These were compositions that were first corrected by a rater, returned to the student, and then re-submitted in revised form. Schmeling then analyzed syntactic differences between the first writing and second writing of only those compositions that had been rated improved.

Sentence Length. A surprising result, and one which challenges one of the most commonly-held intuitions about student writing, is that sentence length may not be as powerful an index for differentiating between "good" and "poor" writers as has been previously thought. Of the 15 factors found to be individually significant, the sentence length index placed just above the cut-off point for non-significance. This finding is not completely surprising, however, in light of the fact that three other investigators, Hunt, Potter, and Schmeling, found sentence length to be a very unreliable indicator of either syntactic maturity or rated quality among native English-speaking students. Thus, with the decidedly more powerful indexes of T-unit length and clause length at our disposal, it would be wise to carefully re-examine any intuitive trust we may have in the discriminating power of the sentence length index.

Passive Voice. Perhaps the most interesting result of this study involves incidence of passive voice. Since it has long been thought that the use of passive voice leads to ineffective writing, it is indeed surprising that there is such a high positive correlation between incidence of passive voice and scores assigned to compositions written by college-level ESL students. Specifically, "good" writers wrote approximately 6½ times as many passives as did "poor" writers, the former using a passive in 25% of their sentences and the latter in only 6% of their sentences. In addition, "good" writers wrote 4 times as many passives per clause, 4 times as many per T-unit, 9 times as many per main clause, and approximately 4½ times as many per secondary clause. The differences, regardless of writing unit, are certainly clear cut. And while all 7 factors dealing with passive voice proved to be significant, all 3 factors involving incidence of active voice were non-significant.

It is surprising that prejudices against the passive continue to flourish, for data from other studies correlate with the findings of this study. Hunt (1965) found that mean incidence of passive voice increased through the grades, correlating at the .01 level of confidence with age. Eighth-graders wrote approximately 3 times as many passives as did 4th-graders, and 12-graders wrote approximately 4 times as many passives as did 4th-graders. Potter (1967) found that "good" 10th-grade writers wrote twice as many passives as did "poor" 10th-grade writers. Similarly, Schmeling (1969) found that among quality-differentiated papers, 'improved' versions of papers contained approximately twice as many passives as did 'first' versions. Finally, Wolk (1969), in his analysis of the writing of 10 professional authors, found that nearly one verb in six was in the passive and that nearly one sentence in three contained at least one passive.

In short, many different types of writers of different ages and language backgrounds use the passive voice quite frequently, with "good" and older writers using more passives than do "poor" and younger writers. Mean incidence of passive voice, then, seems to be a reliable indicator of both syntactic maturity and rated quality of writing, all this suggesting that we should thoroughly re-examine our present attitude toward teaching the passive, for there is simply no apparent justification for our so completely discouraging its use.

Discussion. It is clear that the results of this study have a number of specific and general implications for the ESL composition curriculum.[5] First, in terms of length of writing units, T-unit length and clause length appear to be much more reliable indexes of rated quality than is the time-honored index of sentence length. Statistical counts of these syntactic factors, then, may be used as one part of the overall procedure for properly placing ESL composition students, using these indexes for the same purposes as other objective measures are used in education: "prediction, . . . exemption, growth measurement, (and) program evaluation" (Cooper 1977:16).

Second, the much-maligned passive voice, long considered one of the primary ingredients in the recipe for dull, ineffective writing, apparently does not greatly detract from writing quality, at least in the dialect of formal, expository prose—the type of prose which the assigned topics were designed to elicit, and exactly the type of prose that we are instructing our students to write in when they come to our composition classes. The final answer to the question of *why* quality of writing correlated so highly with incidence of passive cannot be definitively answered here, but perhaps it is that higher incidence of passive voice indicates a greater control over the syntactic structures of the language, in some general sense leaving the reader with the impression that the writer has a higher level of proficiency in the written language. Whatever the answer to this question may be, it is clear that we need to teach our students that, in the dialect of formal, expository prose, the passive is not simply an unattractive transform of the active, but rather a valuable tool in the arsenal of "good" writers.

Third, by our knowing more about what "good" writers put into their writing, and what "poor" writers do not, we will be better able to design exercises to guide "poor" writers in the appropriate direction. Sentence combining exercises would appear to be particularly useful in accomplishing this goal, while exercises that teach the techniques of clause subordination do not appear to be useful.

Finally, it is a solid body of empirical data, not vague intuitions, that will more clearly reveal the nature of the relationship between syntactic skill and ESL writing quality. When we have accumulated this body of experimental data, our job of teaching writing to ESL students, and the students' job of learning to write, may become a great deal less difficult, and perhaps more enjoyable.

---

[5] The limited nature of this study does not imply that syntax is the whole of writing, for such things as content, development, organization, naturalness of expression and audience effectiveness all play a role in determining the quality of a composition.

References

Aitken, Kenneth G. 1977. Using cloze procedure as an overall language proficiency test. *TESOL Quarterly 11*, 1: 59-67.

Alatis, J. E. 1974. Towards a LAPSE theory of teacher preparation in English as a second language. *English Language Teaching 29*, 1:8-18.

Alatis, J. E. 1976a. The past as a prologue. Guest editorial. *TESOL Quarterly 10*, 1:7-18.

Alatis, J. E. 1976b. The compatability of TESOL and bilingual education. In Alatis and Twaddell. (1976), 5-17.

Alatis, J. E. and K. Twaddell (Eds.). 1976. *English as a second language in bilingual education*. Washington, D.C.: TESOL.

Alden, John D. Jan. 22, 1979. Foreign enrollments in engineering. *The Chronicle of Higher Education XVII*, 19.

Alderson, J. C. 1978. The effect of certain methodological variables on cloze test performance and its implication for the use of the cloze procedure in EFL testing. Paper presented at the 5th International Congress of Applied Linguistics, Montreal.

Alderson, J. C. submitted. The effect on the cloze test of changes in deletion frequency.

Allen, H. and R. Campbell (Eds.). 1972. *Teaching English as a second language*. New York: McGraw-Hill.

Allen, J. P. B. 1976. Personal communication cited in Tarone, Swain and Fathman, 1976.

Allen, J. P. B. and H. G. Widdowson. 1978. Teaching the communicative use of English. *English for specific purposes*. Ronald Mackay and Alan Mountford (Eds.). London: Longman, 56-77.

American Personnel and Guidance Association. 1978. APGA policy statement: Responsibilities of users of standardized tests. *Guideposts*. October 5, 1978, 5.

American Psychological Association. 1974. *Standards for educational and psychological tests*. Washington, D.C.: American Psychological Association.

Anastasi, A. and F. A. Córdova. 1953. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. *Journal of Educational Psychology 44*: 1-19.

Andersen, E. 1977. Young children's knowledge of role related speech differences: a mommy is not a daddy is not a baby. Palo Alto, CA: Stanford University, Department of Linguistics.

Anderson, J. 1972. The application of cloze procedure to English learned as a foreign language in Papua and New Guinea. *English Language Teaching XXVII*, 1: 66-72.

Ausubel, D. 1964. Adults versus children in second language learning: psychological considerations. *Modern Language Journal* 48:420-424.

Austin, J. L. 1962. *How to do things with words*. New York: Oxford University Press.

Anthony, E. M. 1963. Approach, method and technique. *English Language Teaching 17*, 63-67.

Bailey, N., C. Madden, and S. Krashen. 1974. Is there a "natural sequence" in adult second language learning? *Language Learning 24*: 235-243.

Bancroft, W. J. 1978. The Lozanov method and its American adaptations. *Modern Language Journal LXII*: 167-175.

Barnes, D. 1976. *From communication to curriculum*. Harmondsworth: Penguin Books.

Bedingfield, Dolores. 1978. The effects of U.S. culture on foreign students. Paper read at South Central Modern Language Association meeting, Houston, Texas.

Bezanson, Keith A. and Nicolas Hawkes. 1976. Bilingual reading skills of primary school children in Ghana. *Working Papers on Bilingualism 11*: 44-73.

Blatchford, C. H. 1976. The silent way and teacher training. *On TESOL '76*. John Fanselow and Ruth Crymes (Eds.). Washington, D.C.: TESOL, 23-28.

Blatchford, C. H. 1977. Some questions on CBTE from TESOL. *Bilingual, ESOL and Foreign Language Teacher Preparation: Models, practices, issues*. John Fanselow and Richard Light (Eds.). Washington, D.C.: TESOL, 191-194.

Blatchford, C. H. (Ed.). 1977. *Directory of teacher preparation programs in TESOL and bilingual education 1976-78*. Washington, D.C.: TESOL.

Blatchford, C. H. (Ed.). 1979a. *MS pages of directory of teacher preparation programs in TESOL and bilingual education: 1978-1980*. Washington, D.C.: TESOL.

Blatchford, C. H. 1979b. Letter and personal communication.

Bolinger, D. 1975. *Aspects of language*. 2d ed. New York: Harcourt Brace Jovanovich.

Bordon, R. and D. Schuster. 1976. The effects of a suggestive learning climate, synchronized breathing and music on the retention of Spanish words. *The Journal of Suggestive-Accelerative Learning and Teaching*, 27-40.

Borkin, Ann. 1978. On some conjuncts signalling dissonance in written expository English. Presented at 1978 TESOL Convention, Mexico City, Mexico.

Bormuth, J. R. 1965. Validities of grammatical and semantic classifications of cloze test scores. *Proceedings of the International Reading Association Conference 10*: 283-286.

Bowen, J. D. 1967. Maximum results from minimum training. *TESOL Quarterly 1*, 2: 23-32.

Bowen, J. D. 1978. The identification of irrelevant lexical distraction: an editing task. *TESL Reporter 12*, 1: 1-3, 14-16.

Bower, T. G. R. 1974. *Development in infancy*. San Francisco: Freeman.

Boyd, John and Mary Ann Boyd. 1978. An overseas view of scientific English. *TESOL Newsletter XII*, 5.

Braddock, Richard, R. Lloyd-Jones, and L. Schoer. 1963. *Research in written composition*. Urbana, IL: National Council of Teachers of English.

Breen, M. P. and C. N. Candlin. Forthcoming. *The communicative curriculum in language teaching*. London: Longman.

Brière, E. J. 1970. Discrete vs. non-discrete testing in English language proficiency. In *English language testing, Report of the RELC fifth regional seminar*. Roy Cherrier (Ed.). Singapore: RELC.

Brière, E. J. 1972. Are we really measuring proficiency with our foreign language tests? *Teaching English as a second language: a book of readings*. H. B. Allen and R. N. Campbell (Eds.). New York: McGraw-Hill, 321-330.

Brod, R. I. 1976. Foreign language enrollments in U.S. colleges—fall 1974. *The Modern Language Journal 60*, 4: 168-171.

Brooks, Nelson. 1966. The ideal preparation of foreign language teachers. *The Modern Language Journal*, Feb. 1966, 71-78.

Brown, H. D., C. A. Yorio, and R. H. Crymes (Eds.). 1977. *On TESOL '77: Teaching and learning English as a second language: Trends in research and practice*. Washington, D.C.: TESOL.

Brown, H. Douglas. 1973. Affective variables in second language acquisition. *Language Learning 23*, 231-244.

Brown, H. Douglas. 1977. Some limitations of C-L/CLL models of second language teaching. *TESOL Quarterly 11*, 365-372.

Brown, P. and S. Levinson. 1974. Universals in linguistic usage: politeness phenomena. Cambridge University, Unpublished manuscript.

Brown, P. 1975. Women are more polite: sex rôles and women's speech in a Mayan community. Cambridge University, Unpublished manuscript.

Brown, P. and S. Levinson. 1978. Universals in language usage: politeness phenomena. *Questions and Politeness: Strategies in social interaction*. E. N. Goody (Ed.). Cambridge: Cambridge University Press.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge: Harvard University Press.

Bruck, Margret and Jeffrey Schultz. 1977. Ethnographic analysis of the language patterns of bilingually schooled children. *Working Papers on Bilingualism 13*, 60-71.

Burt, M., H. Dulay, and M. Finocchiaro (Eds.). 1977. *Viewpoints on English as a second language*. New York: Regents Publishing Company, Inc.

Campbell, Russell N. 1967. On defining the objectives of a short-term training program. *TESOL Quarterly 1*, 4:44.

Candlin, C. N. 1976. Communicative language teaching and the debt to pragmatics, in *Semantics: theory and application*. C. Rameh (Ed.). Georgetown Monograph Series. Washington, D.C.: Georgetown University.

Candlin, C. N. 1977. Preface to M. Coulthard. *An introduction to discourse analysis*. London: Longman.

Carrell, Patricia L. 1977. Empirical investigations of indirectly conveyed meaning: Assertion versus presupposition in first and second language acquisition. *Language Learning 27*, 2:353-369.

Carrell, Patricia L. 1978. The given-new strategy: Testing presupposition and assertion in L_1 and L_2. *On TESOL '78: EFL policies, programs, practices*. C. H. Blatchford and Jacquelyn Schachter (Eds.). Washington, D.C.: TESOL, 232-239.

Carroll, J. B., A. S. Carton, and C. P. Wilds. 1959. An investigation of cloze items in the measurement of achievement in foreign languages. A report on research conducted under a grant from the College Entrance Examination Board. Cambridge, Mass.: Harvard University, Laboratory for Research in Instruction, Graduate School of Education.

Carroll, J. B. 1961. Fundamental considerations in testing for English proficiency of foreign students. *Testing the English proficiency of foreign students*. Washington, D.C.: Center for Applied Linguistics, 31-40.

Carroll, J. B. 1963. Research on teaching foreign languages. *Handbook of research in teaching*. N. L. Gage (Ed.). Chicago, IL: Rand-McNally, 1060-1100.

Carroll, J. B. 1972. Fundamental considerations in testing for English language proficiency of foreign students. *Teaching English as a Second Language*. H. B. Allen and R. N. Campbell (Eds.). New York: McGraw-Hill, 313-321.

Carver, D. and M. J. Wallace. 1975. Some applications of micro-teaching to TESL. *English Language Teaching* 29, 3: 184-190.

Center for Applied Linguistics. 1977. *Bilingual education: Current perspectives*. Vols. 1-5. Arlington, VA: Center for Applied Linguistics.

Chamot, Anna Uhl. 1979. Strategies in acquisition of English by a child bilingual in Spanish and French. In *The acquisition and use of Spanish and English as first and second languages*. Roger Andersen (Ed.). Washington, D.C.: TESOL. In press.

Chastain, K. 1976. *Developing second language skills: Theory to practice. 2d ed.* Chicago: Rand McNally.

Clark, Herbert H. and Susan E. Haviland. 1977. Comprehension and the given-new contract. In *Discourse production and comprehension, Vol. 1, Discourse processes: Advances in research and theory*. Roy O. Freedle (Ed.). Norwood, N.J.: Ablex Publishing Corp., 1-40.

Clarke, M. A. and L. Burdell. 1977. Shades of meaning: Syntactic and semantic parameters of cloze test responses. Paper presented at the TESOL Convention, Miami, Fla.

Cloze, R. A. 1966. What qualifications do we need for the teaching of English as a foreign language? *English Language Teaching 20*, 2: 98-102 and *English Language Teaching 20*, 3: 217-223.

Cohen, Andrew D. 1976. *A sociolinguistic approach to bilingual education*. Rowley, MA: Newbury House Publishers, Inc.

Connolly, H. V. 1972. The oral English proficiency of foreign students at UCLA. Master's thesis, University of California, Los Angeles.

Cooke, M. A. 1978. A pair of instruments for measuring student attitudes toward bearers of the target culture. *Foreign Language Annals 11*, 2: 149-163.

Cooper, Charles R. 1977. Holistic evaluation of writing. *Evaluating Writing*. Charles Cooper and Lee Odell (Eds.). Urbana, IL: National Council of Teachers of English, 3-31.

Combs, A. W. 1972. Some basic concepts for teacher education. *Journal of Teacher Education* 23, 3: 286-290.

Cooper, T. 1976. Measuring written syntactic patterns of second language learners of German. *Journal of Educational Research* 69: 176-183.

Corder, S. Pit. 1976. The study of interlanguage. *Proceedings of the Fourth International Congress of Applied Linguistics.* Gerhard Nickel (Ed.). Stuttgart: Hochschul Verlag, Vol. 2, 9-34.

Corder, S. Pit. 1977. Language teaching and learning: A social encounter. *On TESOL '77: Teaching and learning English as a second language: Trends in research and practice.* H. D. Brown, C. A. Yorio and R. H. Crymes (Eds.). Washington, D.C.: TESOL, 1-13.

Coulthard, Malcolm. 1977. *An introduction to discourse analysis.* London: Longman.

Crymes, Ruth and William Potter. 1979. Questions in the negotiation for understanding. In *English for cross cultural communication.* Larry E. Smith (Ed.). London: Macmillian.

Cummins, J. 1978. Educational implications of mother tongue maintenance in minority-language groups. *The Canadian Modern Language Review*, 34: 395-416.

Cummins, J. 1979a. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research* 49: 222-251.

Cummins, J. 1979b. Bilingual and educational development in anglophone and minority francophone groups in Canada. *Interchange* 9: 40-51.

Curran, Charles A. 1961. Counseling skills adapted to the learning of foreign languages. *Bulletin of the Menninger Clinic* 25, 78-93.

Curran, Charles A. 1972. *Counseling-learning: A whole-person model for education.* New York: Grune and Stratton.

Curran, Charles A. 1976. *Counseling-learning in second languages.* Apple River, IL: Apple River Press.

Cyphert, F. R. 1972. An analysis of research in teacher education. *Journal of Teacher Education* 23, 2: 145-151.

Danoff, M., G. Coles, D. McLaughlin, and D. Reynolds. 1978. *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs, Volume III.* Palo Alto, CA: American Institutes for Research.

Danielson, D. and R. Hayden. 1973. *Using English your second language.* Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Darnell, D. K. 1968. The development of an English language proficiency test of foreign students using a clozentropy procedure. Final Report. DHEW Research Bureau No. BP - 7 - H - 010. Boulder, Colorado: University of Colorado.

Decker, Donald. 1978. The importance of the English language in today's world. *English around the world.* John I. B. McCulloch (Ed.). New York: The English Speaking Union of the United States, Number 19, 1 and 7.

DeLorenzo, W. E. 1975. Micro teaching as a transitional technique to student teaching. *Foreign Language Annals* 8, 3: 239-245.

de Sassure, F. 1959. *A course in general linguistics.* New York: Philosophical Library.

Directory of state foreign language supervisors. 1978. *Foreign Language Annals 11*, 6: 798-710.

Dowling, G. and K. Shepard. 1976. Teacher training: A counseling focus. *On TESOL '76.* J. F. Fanselow and R. H. Crymes (Eds.). Washington, D.C.: TESOL, 55-65.

Dubin, Fraida. 1978. The shape they are in now and the shape of textbooks to come. *On TESOL '78: EFL policies, programs, practices.* Charles H. Blatchford and Jacquelyn Schachter (Eds.). Washington, D.C.: TESOL, 128-133.

Dugan, D. G. 1967. Micro teaching: A promising medium for teacher retraining. *Modern Language Journal 51*, 3: 161-165.

Dulay, H. and M. Burt. 1974. Natural sequence in child second language acquisition. *Language Learning* 23, 245-258.

Dulay, H. and M. Burt. 1977. Remarks on creativity in language acquisition. In *Viewpoints on English as a second language.* M. Burt, H. Dulay, and M. Finocchiaro (Eds.). New York: Regents Publishing Company, Inc., 95-126.

Dunkel, M. 1978. Second language acquisition of register: When does it occur? Paper presented at the second annual Los Angeles Second Language Research Forum.

Educational Testing Service. 1979. *Principles, policies, and procedural guidelines regarding ETS products and services.* Princeton, N.J.: Educational Testing Service.

Edwards, A. L. 1957. *Techniques of attitude scale construction.* New York: Appleton-Century Crofts, Inc.

English: A growing 'foreign language.' 1978. *Chronicle of Higher Education,* May 30, p. 7.

Ervin-Tripp, S. 1974. "The comprehension and production of requests by children," in *Papers and Reports on child language development.* Stanford, CA: Stanford University.

Ewer, J. R. 1975. Teaching English for science and technology: The specialized training of teacher and programme organizers. *English for academic study: Problems and perspectives.* London: English Teaching Information Centre.

Fanselow, J. F. 1977. An approach to competency-based teacher education in second language teaching. *Bilingual, ESOL and foreign language teacher preparation: Models, practices, issues.* J. F. Fanselow and R. L. Light (Eds.). Washington, D.C.: TESOL, 129-139.

Fanselow, J. F. and R. L. Light. 1977. *Bilingual, ESOL and foreign language teacher preparation: Models, practices, issues.* Washington, D.C.: TESOL.

Fanselow, J. F. 1977. Review of Gattegno, *Common Sense of foreign language teaching* and Stevick, *Memory, Meaning and method. TESOL Quarterly 11,* 4: 453-459.

Ferguson, C. A. 1966. Research on language learning. *Report of the Northeast Conference.*

Ferguson, C. A. 1976. The structure and use of politeness formulas. *Sociology* 5, 137-151.

Fillmore, L. W. 1976. The second time around: Cognitive and social strategies in second language acquisition. Unpublished PhD dissertation, Stanford University.

Fillenbaum, S., L. V. Jones, and A. Rapoport. 1963. The predictability of words and their grammatical classes as a function of the rate of deletion from a speech transcript. *Journal of Verbal Learning and Verbal Behaviour 2:* 186-194.

Finocchiaro, M. 1974. *English as a second language: From theory to practice.* New York: Regents Publishing Company.

Finocchiaro, M. and L. Ekstrand. 1977. Migration today. Some social and educational problems. *Viewpoints on English as a second language.* M. Burt, H. Dulay and M. Finocchiaro (Eds.). New York: Regents Publishing Company, 205-218.

Flahive, D. E. 1977. Separating the g factor from reading comprehension. *Occasional papers on Linguistics 1.* Proceedings of the First International Conference on Frontiers in Language Proficiency and Dominance Testing, held at Southern Illinois University at Carbondale, April 21-23, 1977, pp. 53-58. Also in *Research in Language Testing.* J. W. Oller, Jr. and Kyle Perkins (Eds.). Rowley, MA: Newbury House Publishers, Inc., 15-27.

Flick, William C. and Patricia L. Carrell. 1978. Acquisition of indirect requests in ESL. Unpublished paper, presented at TESOL, 1978, Mexico City.

Foreign flood: Those many alien students. 1979. *Time,* January 29, p. 63.

Frank, Marcella. 1972. *Modern English, Parts 1 and 2.* Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Fraser, B. 1975. The concept of politeness. Paper presented at the Fourth Annual Conference on New Ways of Analyzing Variation in English (NWAVE), Georgetown University.

Fries, Charles C. and Agnes C. Fries. 1961. *Foundations for English teaching.* Tokyo: Kenkyush, Ltd.

Forrester, J. 1974. Demonstration lessons. *English Language Teaching* 29, 1:52-54.

Gaies, S. 1977. A comparison of the classroom language of ESL teachers and their speech among peers: an exploratory syntactic analysis. PhD dissertation, Indiana University.

Garcia, Mary H. and Janet Gonzalez-Mena. 1976. *The big E.* Program guide for SR—An English language development program for the primary grades. Silver Spring, MD: Institute of Modern Languages.

Gardner, R. C. and W. E. Lambert. 1972. *Attitudes and motivation in second language learning.* Rowley, MA: Newbury House Publishers, Inc.

Gattegno, Caleb. 1963. *Teaching foreign languages: The silent way.* New York: Educational Solutions, Inc.

Gattegno, Caleb. 1972. *Teaching foreign languages in schools: The silent way. 2nd ed.* New York: Educational Solutions, Inc.

Gattegno, Caleb. 1976. *The common sense of foreign language teaching.* New York: Educational Solutions, Inc.

Goffman, E. 1967. *Interaction ritual.* Garden City, New York: Doubleday and Company, Inc.

Goffman, E. 1972. *Interaction ritual.* Harmondsworth, Allen Lane: The Penguin Press.

Graham, Carolyn. 1979. *Jazz chants for children.* New York: Oxford University Press.

Grice, H. Paul. 1975. Logic and conversation. In *Syntax and semantics, Vol. 3, Speech acts.* Peter Cole and Jerry Morgan (Eds.). New York: Academic Press, Inc., 41-58.

Grimshaw, Allen D. 1973. Rules, social interaction, and language behavior. *TESOL Quarterly* 7, 2: 99-115.

Guidelines for the certification and preparation of teachers of English to speakers of other languages in the United States. 1975. Washington, D.C.: TESOL.

Guidelines for the preparation of teachers of English—an exposition. English teacher preparation study. 1968. *English Journal* 57, 4.

Guidelines for teacher education programs in modern foreign languages—an exposition. 1966. *Modern Language Journal 50*, 6: 323-425.

Gumperz, J. J. 1964. Linguistic and social interaction in two communities. In *American Anthropologist.* J. J. Gumperz and D. Hymes (Eds.). 66, 6: 2.

Gumperz, J. J. and D. Hymes (Eds.). 1972. *Directions in sociolinguistics: The ethnography of communication.* New York: Holt, Rinehart and Winston.

Hafner, L. E. 1964. Relationships of various measures to the cloze. *13th Yearbook of the National Reading Conference*, 135-145.

Hakuta, K. and H. Cancino. 1977. Trends in Second Language Acquisition Research. *Harvard Educational Review 47*, 3: 294-316.

Halliday, M., A. McIntosh, and P. Strevens. 1970. *The linguistic sciences and language teaching.* London: Longman.

Halliday, M. 1973. The functional basis of language. In *Class, codes and control, Vol. 2: Applied studies towards a sociology of language.* B. Bernstein (Ed.). London: Routledge & Kegan Paul.

Hanchey, Stephen. 1976. Peace Corps language training—where we stand. Unpublished paper.

Hatch, E. 1978. Discourse analysis and second language acquisition. *Second language acquisition: A book of readings.* E. Hatch (Ed.). Rowley, MA: Newbury House Publishers, 401-435.

Hébert, R. et al. 1976. Rendement académique et langue d'enseignement chez les élèvesnent franco-manitobains. Saint-Boniface, Manitoba: Centre de recherches du Collège Universitaire de Saint-Boniface.

Hepworth, George R. 1978. A functional description of the discourse for the English of science and technology. Unpublished MA Thesis. Logan, Utah: Utah State University.

Higa, Masanori. 1965. The psycholinguistic concept of 'difficulty' and the teaching of foreign language vocabulary. *Language Learning 15*, 4: 167-179.

Hinofotis, Frances. B. 1976. An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language. Unpublished PhD dissertation, Southern Illinois University at Carbondale.

Hinofotis, Frances. B. 1977. Cloze testing as a substitute for oral interviews. Paper presented at the pre-convention workshop on cloze testing, TESOL conference, Miami.

Hirsch, E. D., Jr. 1977. *The philosophy of composition.* Chicago: University of Chicago Press.

Hobbs, Jerry R. and Jane J. Robinson. 1978. Why ask? Unpublished paper. (Technical Note 169, SRI International, Menlo Park, California).

Houp, Kenneth W. and Thomas E. Pearsall. 1977. *Reporting technical information.* Beverly Hills, California: Glencoe Press.

Householder, F. 1971. The primacy of writing. *Linguistic Speculations.* F. Householder. Cambridge: Cambridge University Press, 244-264.

Hunt, Kellogg W. 1965. *Grammatical structures written at three grade levels,* Research Report No. 3. Urbana, IL: National Council of Teachers of English.

Hunt, Kellogg W. 1977. Early blooming and late blooming syntactic structures. *Evaluating writing.* Charles Cooper and Lee Odell (Eds.). Urbana, IL: National Council of Teachers of English, 91-104.

Hursh, D. E. 1976. Personalized systems of instruction: What do the data indicate? *Journal of Personalized Instruction 1.*

Hymes, D. 1967. Models of the interaction of language and social setting. *Journal of Social Issues* 23, 2: 8-28.

Hymes, D. 1968. The ethnography of speaking. In *Readings in the sociology of language.* J. A. Fishman (Ed.). The Hague: Mouton.

Hymes, D. 1972. Models of the interaction of language and social life. *Directions in sociolinguistics.* J. J. Gumperz and D. Hymes (Eds.). New York: Holt, Rinehart and Winston, 35-71.

Hymes, D. 1974. *Foundations in sociolinguistics.* Philadelphia: University of Pennsylvania Press.

IIE Reports. 1978. A communication from the Institute of International Education. 3, 1. New York: Institute of International Education.

235,509 foreign students on U.S. campuses last year: Total doubles in last ten years. 1979. *IIE Reports 4,* 1: 5-8.

Ingram, E. 1964. English language battery (ELBA). Edinburgh: University of Edinburgh, Department of Linguistics.

Ingram, E. 1973. English standards for foreign students. *University of Edinburgh Bulletin* 9, 12: 4-5.

Jakobovits, L. A. 1970. A fundamental approach to the assessment of language skills. *Foreign language learning: A psycholinguistic analysis of the issues.* L. A. Jakobovits (Ed.). Rowley, MA: Newbury House Publishers, Inc., 150-165.

Jarvis, G. A. 1972. Teacher education goals: They're tearing up the street where I was born. *Foreign Language Annals* 6, 2: 198-205.

Jensen, Arthur R. 1969. How much can we boost IQ and scholastic achievement? *Harvard Educational Review* 39: 1-123.

Johansson, Stig. 1973. Partial dictation as a test of foreign language proficiency. *Swedish-English contrastive studies report no.* 3. Sweden: Lund University, Department of English.

Johnson, P. O. and J. Neyman. 1936. Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs 1,* 57-93.

Johnson, K. R. 1977. Proctoring training for natural control. *Journal of personalized Instruction* 2, 4.

Johnson, Keith. 1977. Adult beginners: A functional or just a communicative approach? Paper delivered at the B.A.A.L. Annual Meeting, Colchester.

Jones, K. and P. Roe. 1975. Designing English for science and technology (EST) programmes. *English for academic study.* London: English Teaching Information Centre.

Katz, J. J. 1972. *Semantic theory.* New York: Harper and Row.

Kameen, Patrick T. 1978. A mechanical, meaningful, and communicative framework for ESL sentence combining exercises. *TESOL Quarterly 12,* 395-401.

Kerlinger, F. 1973. *Foundations of behavioral research.* New York: Holt, Rinehart and Winston.

Krashen, S. D. 1977. Some issues relating to the monitor model. *On TESOL '77: Teaching and learning English as a second language: Trends in research and* *practice*. H. D. Brown, C. A. Yorio, and R. H. Crymes (Eds.). Washington, D.C.: TESOL, 144-158.

Krohn, Robert. 1971. *English Sentence structure*. Ann Arbor, Michigan: University of Michigan Press.

Kulik, J. A., C-L. Kulik, and B. B. Smith. 1976. Research on personalized system of instruction. *Programmed Learning and Educational Technology 13*.

Kunkle, J. F. 1977. The 'L' connection: The interdependence of second language programs. In Phillips (1977), 211-231.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: Univ. of Pennsylvania Press.

Lackstrom, John E., Larry Selinker and Louis P. Trimble. 1970. Grammar and technical English. *English as a second language: Current issues*. Robert C. Lugton (Ed.). Philadelphia: The Center for Curriculum Development, 101-103.

Lackstrom, John E., Larry Selinker and Louis P. Trimble. 1975. The reading comprehension of elliptical arguments in EST textbooks. Paper presented at the EILI meetings, Stuttgart, Germany.

Lado, Robert. 1957. *Linguistics across cultures: applied linguistics for language teachers*. Ann Arbor, Michigan: University of Michigan Press.

Lado, Robert. 1961. *Language testing: The construction and use of foreign language tests*. New York: McGraw-Hill.

LaFay, H., James L. Stanfield and Louis S. Glanzman. 1978. EBLA: splendor of an unknown empire. *National Geographic 154*, 730-759.

LaForge, Paul G. 1971. Community-language learning: A pilot study. *Language Learning 21*, 45-61.

Lakoff, Robin. 1973. The logic of politeness; or minding your p's and q's. *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*. Chicago: University of Chicago, Chicago Linguistic Society, 292-305.

Lambert, W. E. and G. Richard Tucker. 1972. *Bilingual education for children*. Rowley, MA: Newbury House Publishers, Inc.

Lambert, W. E. 1975. Culture and language as factors in learning and education. *Education of Immigrants*. A. Wolfgang (Ed.). Toronto: The Ontario Institute for Studies in Education, 55-83.

Lamendella, John T. 1979. The neurofunctional basis of pattern practice. *TESOL Quarterly 13*, 1: 5-19.

Larsen-Freeman, Diane. 1974. A re-evaluation of grammatical structure sequencing. *On TESOL 1974*. W. Norris and R. H. Crymes (Eds.). Washington, D.C.: TESOL, 151-161.

Larsen-Freeman, Diane. 1975. The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly* 9, 409-419.

Larsen-Freeman, Diane and V. Strom. 1977. The construction of a second language acquisition index of development. *Language Learning 27*, 123-134.

Larsen-Freeman, Diane. 1978. An ESL index of development. *TESOL Quarterly 12*, 4: 439-448.

Lee, R. R. 1972. A taxonomy of second language behaviors. *TESOL Quarterly 6*, 3: 209-220.

Lee, W. R. 1974. On getting down to grass roots in EFL teacher training. *English Language Teaching* 29, 1: 35-44.

Lee, W. R. 1977. What type of syllabus for the teaching of English as a foreign or second language? *IRAL 15*, 8: 246-249.

Leech, G. 1974. *Semantics*. Harmondsworth: Penguin Books.

Lett, J. A. 1977. Assessing attitudinal outcomes. In Phillips (1977), 265-267.

Littlewood, W. T. 1975. Role-performance and language teaching. *IRAL 13*, 3.

Loban, W. 1976. *Language development: kindergarten through grade twelve*. Research Report no. 18. Urbana, IL: National Council of Teachers of English.

Logan, Lillian M., Virgil G. Logan and Leonoa Paterson. 1972. *Creative communication: Teaching the language arts*. New York: McGraw-Hill.

Longacre, R. E. 1970. Sentence structure as a statement calculus. *Language 46*, 4: 783-815.

Longacre, R. E. 1976. A taxonomic deep and surface structure analysis of 'The lover and his lass'. Summer Institute of Linguistics, August 24, 1976.

Lord and T'sou (Eds.). 1976. *Studies in bilingual education: Selected papers from the 1976 International symposium on bilingual education, Hong Kong.* Hong Kong: Lammar Offset Printing.

Lozanov, G. 1978. *Suggestology and outlines of suggestopedy.* New York: Gordon and Breach Science Publishers, Inc.

Mackey, William F. 1965. *Language teaching analysis.* London: Longman.

Mackey, William F. 1972. *Bilingual education in a binational school.* Rowley, MA: Newbury House Publishers, Inc.

Mackay, R. 1978. Identifying the nature of the learner's needs. *English for specific purposes.* R. Mackay and A. Mountford (Eds.). London: Longman, 21-37.

Madsen, Harold S. and J. Donald Bowen. 1978. Adapting ESL teaching materials. *On TESOL '78: EFL Policies, Programs, Practices.* Charles H. Blatchford and Jacquelyn Schachter (Eds.). Washington, D.C.: TESOL, 137-148.

Marqwardt, M. F. 1969. Preparing English teachers abroad. *TESOL Quarterly* 3, 1: 31-36.

Martin-Jones, M. 1978. A second language in contest: new perspectives on syllabus design and teaching. Unpublished M.A. thesis, Linguistics Department, Stanford Univ.

McArdle, R. T. 1971. Training language teachers—a systems approach. *Foreign language Annals 4*, 3: 293-298.

McClintock, Robert. 1971. Toward a place for study in a world of instruction. *Teachers College Record* 73, 2. New York: Columbia University.

Mellgren, Lars and Michael Walker. 1979. *Young English Series: YES!* Book D Teachers' Guide. Reading, MA: Addison-Wesley Publishing Company.

Miele, P. 1978. The power of suggestion—a new way of learning languages. *Parade* May: 14-15.

Mitchell-Kernan, C. and K. Kernan. 1977. Pragmatics of discourse choice among children. *Child Discourse.* C. Mitchell-Kernan and S. Ervin-Tripp (Eds.). New York: Academic Press.

Monroe, J. 1975. Measuring and enhancing syntactic fluency in French. *The French Review* 48, 1023-1031.

Montaigne. Of training. In Montaigne, *Selected essays.* C. Cotton and W. Hazlitt (Trans.). New York: The Modern Library, 116-128.

Moore, A. C. 1977. Teacher training: The demonstration lesson. *English Language Teaching 31*, 3: 208-211.

Morgan, Jerry L. 1978. Two types of convention in indirect speech acts. *Syntax and Semantics Vol. 9, Pragmatics.* Peter Cole (Ed.). New York: Academic Press, 261-280.

Moskowitz, G. 1976. Competency-based teacher education: Before we proceed. *Modern Language Journal 60*, 1 & 2: 18-23.

Mountain, Lee Harrison. 1974. *The do-it-yourself guide to teaching language arts.* Highland Park, New York: Dreier Educational Systems.

Mullen, K. A. 1977a. Rater reliability and oral proficiency evaluations. In *Proceedings of the first international conference on frontiers in language proficiency and dominance testing. Occasional papers on linguistics 1.* J. E. Redden (Ed.). Carbondale, Ill.: Department of Linguistics, Southern Illinois University, 133-42.

———. 1977b. Using rater judgments in the evaluation of writing proficiency for non-native speakers of English. In *Teaching and learning English as a second language: trends in research and practice.* H. D. Brown, C. A. Yorio and R. H. Crymes (Eds.). Washington, D.C.: TESOL, 309-20.

———. 1978. Determining the effect of uncontrolled sources of error in a direct test of oral proficiency and the capability of the procedure to detect improvement following classroom instruction. In *Direct testing of speaking proficiency: theory and application.* J. L. D. Clark (Ed.). Princeton, N.J.: Educational Testing Service, 171-88.

Mullen, Karen. 1979a. Evaluating writing proficiency in ESL. In Oller and Perkins (1979), 127-134.

Mullen, Karen. 1979b. Rater reliability and oral proficiency evaluations. Revised version of a paper from *Occasional Papers on Linguistics No. 1.* Proceedings of the First International Conference on Frontiers in Language Proficiency and

Dominance Testing, held at Southern Illinois University at Carbondale, April 21-23, 1977.

Mullen, Karen. 1979c. More on cloze tests as tests of proficiency in English as a second language. *Concepts in language testing: Some recent studies*. E. J. Brière and F. B. Hinofotis (Eds.). Washington, D.C.: TESOL, 21-32.

Munby, J. 1977. Designing a processing model for specifying communicative competence in a foreign language: A study of the relationship between communicative needs and the English required for special purposes. Unpublished doctoral thesis, University of Essex. (forthcoming).

Naiman, N., M. Fröhlich, and H. H. Stern. 1975. *The good language learner*. Toronto: Modern Language Centre, The Institute for Studies in Education.

Nakano, P. J. 1977. Educational implications of the *Lau v. Nichols* decision. *Viewpoints on English as a second language*. M. Burt, H. Dulay and M. Finocchiaro (Eds.). New York: Regents Publishing Company, 219-234.

Nanba, K., E. Saban, M. C. Bowling, and C. W. Bernard. 1977. Splenic pseudosinuses and hepatic angiomatous lesions. *American Journal of Clinical Pathology* 67, 5: 415-526.

Neisser, U. 1976. *Cognition and reality*. San Francisco: W. H. Freeman and Co.

Nunnally, J. C. 1967. *Psychometric theory*. New York: McGraw-Hill.

O'Donnell, R., W. Griffin, and R. C. Norris. 1967. Syntax of kindergarten and elementary school children: A transformational analysis. Research report no. 8. Urbana, IL: National Council of Teachers of English.

Oller, J. 1970. Dictation as a device for testing foreign language proficiency. *Workpapers in TESL: UCLA* 4: 37-42. Also in *English Language Teaching* 25: 254-259.

Oller, J. and C. Conrad. 1971. The cloze procedure and ESL proficiency. *Language Learning* 21: 183-196.

Oller, J. 1972a. Dictation as a test of English language proficiency. *Teaching English as a second language*. H. B. Allen and R. N. Campbell (Eds.). New York: McGraw-Hill, 346-354.

Oller, J. 1972b. Scoring methods and difficulty levels for cloze tests of proficiency in ESL. *Modern Language Journal* 56: 151-158.

Oller, J. 1973a. Discrete point tests versus tests of integrative skills. *Focus on the learner*. J. Oller and J. Richards (Eds.). Rowley, MA: Newbury House Publishers, 184-199.

Oller, J. 1973b. Some psycholinguistic controversies. *Focus on the learner*. J. Oller and J. Richards (Eds.). Rowley, MA: Newbury House Publishers, 36-50.

Oller, J. and J. Richards (Eds.). 1973. *Focus on the learner*. Rowley, MA: Newbury House Publishers.

Oller, J., P. Irvine and P. Atai. 1974. Cloze, dictation, and the test of English as a foreign language. *Language Learning* 24, 2: 245-252.

Oller, J. and V. Streiff. 1975. Dictation: a test of grammar based expectancies. *English Language Teaching* 30: 25-36. Also in *Testing language proficiency*. R. Jones and B. Spolsky (Eds.). Arlington, VA: Center for Applied Linguistics, 71-88.

Oller, J. 1976a. Evidence for a general language proficiency factor: an expectancy grammar. *Die Neuren Sprachen* 76: 165-174.

Oller, J. 1976b. Language testing. *Survey of applied linguistics*. R. Wardhaugh and H. D. Brown (Eds.). Ann Arbor: University of Michigan Press, 275-300.

Oller, J. and F. Hinofotis. 1976. Two mutually exclusive hypotheses about second language ability: Factor analytic studies of a variety of language tests. Paper delivered at winter meeting of Linguistics Society of America, December 30, 1976. Also in *Research in language testing*. J. Oller and K. Perkins (Eds.). Rowley, MA: Newbury House Publishers, 1979, 2-8.

Oller, J. and K. Perkins (Eds.). 1978. *Language in education: Testing the tests*. Rowley, MA: Newbury House Publishers.

Oller, J. 1979a. Competency testing and the language factor. Paper presented at the Thirteenth Annual Meeting of TESOL, Boston, MA. Also in Streiff and Oller (forthcoming).

Oller, J. 1979b. *Language tests at school*. London: Longman.

Oller, J. and K. Perkins (Eds.). 1979. *Research in language testing*. Rowley, MA: Newbury House Publishers.

Oller, J. and S. Chesarek. Forthcoming. Achievement, non-verbal intelligence, and language: a bilingual case study. In Streiff and Oller (forthcoming).

Orwen, G. P. 1971. Blueprint for a methods course. *Modern Language Journal 55*, 7: 456-459.

Ostrander, S. and L. Schroeder. 1970. Psychic discoveries behind the iron curtain. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Pack, A. C. 1973. Cloze testing and procedure. *TESL Reporter 6*, 2: 1-2.

Paulston, C. B. 1973. A biased bibliography: Comments on selecting texts for a methods course in TESOL. *Language Learning* 23, 1: 129-143.

Paulston, C. B. 1974. Linguistic and communicative competence. *TESOL Quarterly* 8, 4: 347-362.

Paulston, C. B. and M. N. Bruder. 1976. *Teaching English as a second language: Techniques and procedures*. Cambridge, MA: Winthrop Publishers.

Phillips, Jane K. 1977. The language connection: From the classroom to the world. Skokie, IL: National Textbook Company.

Pitkin, Willis L., Jr. 1969. Discourse blocs, *College Composition and Communication*. XX, No. 2, 138-147.

———. 1973. *A Pedagogical model of discourse structure*. Unpublished doctoral dissertation.

———. 1977a. Hierarchies and the discourse hierarchy, *College English* 38, 7: 648-659.

———. 1977b. X/Y: Some basic strategies of discourse, *College English* 38, 7: 660-672.

Plato. Republic, from *The Republic of Plato*, F. M. Cornford (Tran.). New York: Oxford University Press, 1945.

Politzer, R. L. 1966. Toward a practice-centered program for the training and evaluation of foreign language teaching. *Modern Language Journal 50*, 5: 251-255.

Potter, Robert R. 1967. Sentence structure and prose quality: An exploratory study. *Research in the Teaching of English 1*, 1: 17-28.

Racle, G. L. 1979. Can suggestopaedia revolutionize language teaching? *Foreign Language Annals 12*, 40.

Rand, E. 1972. Integrative and discrete-point tests at UCLA. *Workpapers in TESL*. University of California, Los Angeles.

Rankin, E. F. 1957. An evaluation of the cloze procedure as a technique for measuring reading comprehension. PhD dissertation, University of Michigan.

Rardin, Jennybelle. 1975. A community language learning approach to a Peace Corps language training program in Afghanistan. Unpublished paper.

Rardin, Jennybelle. 1976. A model for second language learning. In Curran, Charles A., *Counseling-learning in second languages*.

Resolution on Competency-Based Teacher Education (NCTE and TESOL). In Fanselow and Light 1977.

Richterich, R. 1973. Definition of language needs and types of adults. Strasbourg: Council of Europe.

Rivers, Wilga M. 1964. *The psychologist and the foreign-language teacher*. Chicago: University of Chicago Press.

Rivers, Wilga M. 1968. *Teaching foreign-language skills*. Chicago: University of Chicago Press.

Rivers, Wilga M. 1972. *Speaking in many tongues*. Rowley, MA: Newbury House Publishers, Inc.

Rivers, Wilga M. 1976. *Speaking in many tongues*. Exp. 2d ed. Rowley, MA: Newbury House Publishers.

Rivers, Wilga M. and Mary S. Temperley. 1978. *A practical guide to the teaching of English as a second or foreign language*. New York: Oxford University Press.

Rogers, Carl A. 1965. *Client-centered therapy: Its current practice, implications, and theory*. Boston: Houghton-Mifflin Co.

Rosier, P. and M. Farella. 1976. Bilingual education at Rock Point: Some early results. *TESOL Quarterly 10*, 4: 379-388.

Rubin, Dorothy. 1975. *Teaching elementary language arts.* New York: Holt, Rinehart and Winston.

Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. *A simplest systematics for the organization of turn-taking for conversation. Language* 50, 696-735.

Salus, P. (Ed.). 1969. *On Language: Plato to von Humboldt.* New York: Holt, Rinehart and Winston.

Sanneh, O. 1977. A computer analysis of the English as a second language placement examination (ESLPE) results, 1974, '75 and '76. Unpublished paper, Office of International Students and Scholars, University of California, Los Angeles.

Sapir, E. 1921. *Language.* New York: Harcourt, Brace and World.

Savignon, Sandra J. 1972. *Communicative competence: An experiment in foreign language teaching.* Philadelphia: Center for Curriculum Development.

Schmeling, Herman Harold. 1969. A study of the relationship between certain syntactic features and overall quality of college freshman writing. Unpublished PhD dissertation, George Peabody College for Teachers.

Schmieder, A. A., J. L. Mark, and J. L. Aldrich. 1977. Competency-based education: Some probable questions and some possible answers. In Fanselow and Light 1977, 61-64.

Schmidt, R. 1978. Getting what you want in a new language: the use, comprehension, and acquisition of English directives by non-native speakers. Paper presented at the 1978 Los Angeles Second Language Research Forum.

Scholz, G., Debby Hendricks, Randon Spurling, Marianne Johnson, and Lela Vandenburg. 1979. Is language ability divisible or unitary? In Oller and Perkins (1979) 9-14.

Scott, M. and G. R. Tucker. 1974. Error analysis and English-language strategies of Arab students. *Language Learning 24,* 1: 69-97.

Scovel, Thomas. 1978. The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning* 28, 129-142.

Scovel, Thomas. 1979. Review of Lozanov. *TESOL Quarterly* 13, 2: 255-266.

Searle, John R. 1969. *Speech acts.* Cambridge: Cambridge University Press.

Searle, John R. 1975. Indirect speech acts. In *Syntax and semantics, Vol.* 3, *Speech acts,* Peter Cole and Jerry Morgan (Eds.). New York: Academic Press, Inc., 59-82.

Selinker, L. and L. Trimble. 1974. Formal written communication and ESL. *Journal of Technical Writing and Communication* 4, 2.

Shimanoff, S. 1977. Investigating politeness. *Discourse across time and space: Occasional papers in linguistics* No. 5. E. O. Keenan and T. Bennett (Eds.). Los Angeles: University of Southern California, Department of Linguistics.

Silva, Carmen. 1975. Recent theories of language acquisition in relation to a semantic approach in foreign-language teaching. *English Language Teaching Journal* 29, 4: 337-346.

Silverman, R. J., J. K. Noa, and R. H. Russell. 1976. *Oral language tests for bilingual students: An evaluation of language dominance and proficiency instruments.* Portland, OR: Northwest Regional Educational Laboratory.

Simon, S. B., L. M. Howe, and H. Kirschenbaum. 1972. *Values clarification: A handbook of practical strategies for teachers and students.* New York: Hart.

Ekuttnabb-Kangas, T. and P. Toukomaa. 1979. Semilingualism and middle class bias: A reply to Cora Brent-Palmer. *Working Papers on Bilingualism 19.* In press.

Smith, F. 1975. The relation between spoken and written language. *Foundation of language development: A multi-disciplinary approach II.* H. Lenneberg and E. Lenneberg (Eds.). New York: Academic Press, Inc.

Smith, Larry (Ed.). 1978. *English for cross-cultural communication.* Honolulu, Hawaii: East-West Center.

Spearman, C. 1904. "General intelligence" objectively determined and measured. *American Journal of Psychology* 15: 201-293.

Spearman, C. 1927. *The abilities of man.* New York: Macmillan.

Spearman, C. 1950. *Human ability.* London: Macmillan.

Spolsky, Bernard (Ed.). 1972. *The language education of minority children.* Rowley, MA: Newbury House Publishers, Inc.

Spolsky, Bernard. 1973. What does it mean to know a language, or how do you get someone to perform his competence? *Focus on the learner*. J. Oller and J. Richards (Eds.). Rowley, MA: Newbury House Publishers, Inc., 164-176.

Spolsky, Bernard and Robert Cooper (Eds.). 1977. *Frontiers of bilingual education*. Rowley, MA: Newbury House Publishers, Inc.

Stevick, Earl W. 1971. *Adapting and writing language lessons*. Washington, D.C.: FSI, Department of State.

Stevick, Earl W. 1972a. Technemes and the rhythm of class activity. In Allen and Campbell (1972), 88-94.

Stevick, Earl W. 1972b. Evaluating and adapting language materials. In Allen and Campbell (1972), 101-120.

Stevick, Earl W. 1973. Review of Curran, Counseling-learning: A whole-person model for education. *Language Learning* 23, 259-271.

Stevick, Earl W. 1976. *Memory, meaning and method*. Rowley, MA: Newbury House Publishers, Inc.

Streiff, Paul. Forthcoming. The language factor in the CTBS battery with a bilingual population. In Streiff and Oller.

Streiff, Virginia. 1978. Relationship among oral and written cloze scores and achievement test scores in a bilingual setting. In Oller and Perkins (1978), 65-102.

Streiff, Virginia. 1979. Another look at the English language factor in educational tests for bilinguals. Paper presented at the Thirteenth Annual Meeting of TESOL, Boston, Mass. Also in Streiff and Oller (forthcoming).

Strevens, P. 1974. Some basic principles of teacher training. *English Language Teaching* 29, 1: 19-27.

Strevens, Peter. 1977a. Causes of failure and conditions for success in the learning and teaching of foreign languages. In Brown, Yorio and Crymes (1977), 266-277.

Strevens, Peter. 1977b. *New orientations in the teaching of English*. Oxford: Oxford University Press.

Strevens, Peter. 1977c. English for special purposes: An analysis and survey. *Studies in language learning*. J. Ronayne Cowan (Ed.). Urbana, IL: University of Illinois, 111-135.

Strevens, Peter. Forthcoming. English for international and intranational purposes: A shift in linguistic perspectives.

Stubbs, J. B. and G. R. Tucker. 1974. The cloze test as a measure of English proficiency. *Modern Language Journal* 58: 239-241.

Stump, T. 1978. Cloze and dictation tasks as predictors of intelligence and achievement scores. In Oller and Perkins (1978), 36-63.

Swain, Merrill. 1976. Bibliography: Research on immersion education for the majority child. *The Canadian Modern Language Review* 32, 592-596.

Swain, Merrill. 1978. Bilingual education for the English-Canadian. *Georgetown University Round Table on Languages and Linguistics 1978*. J. E. Alatis (Ed.), Washington, D.C.: Georgetown University, 141-154.

Swain, Merrill. 1979. Target language use in the wider environment as a factor in its acquisition. Paper presented at the 13th annual TESOL Convention, March, 1979.

Swain, M., G. Dumas, and N. Naiman. 1974. Alternatives to spontaneous speech: elicited imitation and translation as indicators of second language competence. *Working Papers on Bilingualism* 3: 68-79.

Swain, M. and M. Bruck. 1976. Proceedings of the research conference on immersion education for the majority child. *The Canadian Modern Language Review*. Special issue 32, entire no. 5.

Swain, M., S. Lapkin, and H. Barik. 1976. The cloze test as a measure of second language proficiency for young children. *Working Papers on Bilingualism* 11: 32-44.

Swain, Merrill and J. Cummins. 1979. Bilingualism, cognitive functioning and education. *Language Teaching and Linguistics: Abstracts*, 4-18.

Tarone, E., M. Swain, and Ann Fathman. 1976. Some limitations to the classroom applications of current second language acquisition research. *TESOL Quarterly* 10, 1: 19-32.

Task force findings specifying remedies available for elementary past educational practices ruled unlawful under Lau v. Nichols. Office for Civil Rights. 1975. Washington, D.C.: Department of Health, Education and Welfare.

Taylor, B. P. 1978. Teaching teachers to teach: Tips and techniques. Paper presented at the 12th Annual TESOL Convention, Mexico City.

Taylor, B. P. and N. Wolfson. 1978. Breaking down the free conversation myth. *TESOL Quarterly 12*, 1: 31-39.

Taylor, W. L. 1953. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly 30*, 415-433.

Thurstone, L. L. and T. G. Thurstone. 1941. *Factorial studies of intelligence.* Chicago: University of Chicago.

Trachtenburg, S. 1979. Joke-telling as a tool in ESL. *TESOL Quarterly 13*, 1: 89-99.

Trager, G. L. 1949. The field of linguistics. *SIL Occasional Papers*, Vol. 1. Oklahoma City: S.I.L.

Troike, R. 1978. *Research evidence for the effectiveness of bilingual education.* Los Angeles: National Dissemination and Assessment Center, California State University.

Tucker, G. R. 1977. Can a second language be taught? *On TESOL '77: Teaching and learning English as a second language.* H. D. Brown, C. A. Yoria, and R. Crymes (Eds.). Washington, D.C.: TESOL, 14-30.

Upshur, john A. 1969. TEST is a four-letter word. Paper presented to the EPDA Institute, University of Illinois.

Vachek, J. 1973. *Written language: General problems and problems of English.* The Hague: Mouton.

van Ek, J. A. 1977. *The threshold level for modern language learning in schools.* London: Longman.

Vygotsky, Lev. 1934. *Thought and language.* Tr. Eugenia Haufmann and Gertrude Vakar. Cambridge: MIT, 1962.

Waggoner, D. 1978. Non-English language background persons: Three U.S. surveys. *TESOL Quarterly 12*, 3: 247-262.

Watkins, B. Dec. 11, 1978. Foreigners' enrollment concerns graduate deans. *The Chronicle of Higher Education XVII*, 15.

Weisberg, R. and S. Buker. 1978. Strategies for teaching the rhetoric of written English for science and technology. *TESOL Quarterly 12*, 3: 321-329.

Whitaker, S. F. 1976. What is the status of dictation? *Audio Visual Journal* 14: 87-93.

Widdowson, H. G. 1978. *Teaching language as communication.* London: Oxford University Press.

Wilkins, D. A. 1972. An investigation into the linguistic and situational content of the common core in a unit/credit system, (ms.), Strasbourg: Council of Europe.

Wilkens, D. A. 1974. Notional syllabuses and the concept of a minimum adequate grammar. In *Linguistic insights in applied linguistics.* S. P. Corder and E. Roulet (Eds.). Brussels: AIMAV, Paris: Didier.

Wilkins, D. A. 1976. National Syllabuses. London: Oxford University Press.

Wolk, A. 1969. The passive mystique: We've been had. *English Journal* 58, 432-435.

Wolpert, L. 1978. Pattern formation in biological development. *Scientific American* 239: 154-164.

Yorkey, R. C., R. Barrutia, A. Chamot, et al. 1978. *English for International Communication*, Book 3. New York: Litton Educational Publishing International.

Yorozuya, R. and J. Oller. In press. Construct validity of oral proficiency scales. *Language Learning.*

Zappert, L. and B. R. Cruz. 1977. Bilingual education: An appraisal of empirical research. Berkeley: Bay Area Bilingual Education League/Lau Center, Berkeley Unified School District.